Conditional Mixture Path Guiding for Differentiable Rendering

ZHIMIN FAN, State Key Lab for Novel Software Technology, Nanjing University, China
PENGCHENG SHI, State Key Lab for Novel Software Technology, Nanjing University, China
MUFAN GUO, State Key Lab for Novel Software Technology, Nanjing University, China
RUOYU FU, State Key Lab for Novel Software Technology, Nanjing University, China
YANWEN GUO, State Key Lab for Novel Software Technology, Nanjing University, China
JIE GUO*, State Key Lab for Novel Software Technology, Nanjing University, China

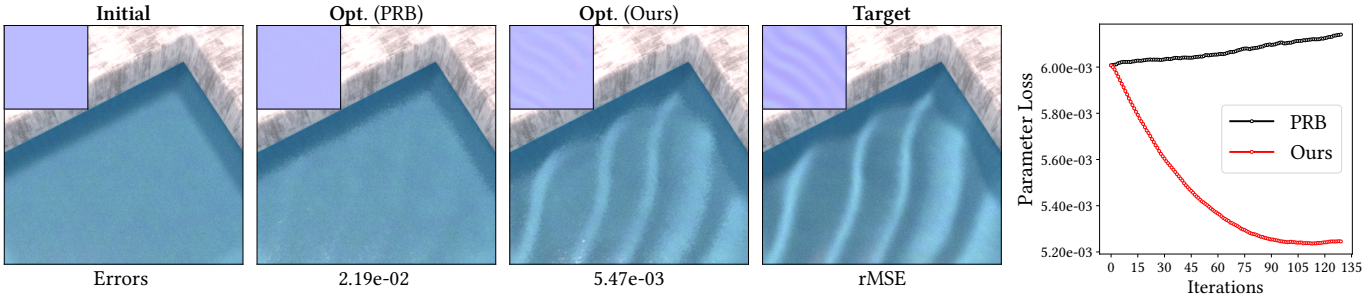

Fig. 1. Reconstructing the normal map of the water surface (roughness $\alpha = 0.1$) from the caustics, using our physically based differentiable rendering pipeline. Given target images and initial scene parameters, existing methods (e.g., PRB [Vicini et al. 2021]) fail due to the challenging specular-diffuse-specular (SDS) paths in the scene, while our method works well. The last column shows the evolution of the parameter loss with respect to the optimization iteration.

The efficiency of inverse optimization in physically based differentiable rendering heavily depends on the variance of Monte Carlo estimation. Despite recent advancements emphasizing the necessity of tailored differential sampling strategies, the general approaches remain unexplored.

In this paper, we investigate the interplay between local sampling decisions and the estimation of light path derivatives. Considering that modern differentiable rendering algorithms share the same path for estimating differential radiance and ordinary radiance, we demonstrate that conventional guiding approaches, conditioned solely on the last vertex, cannot attain this density. Instead, a mixture of different sampling distributions is required, where the weights are conditioned on all the previously sampled vertices in the path. To embody our theory, we implement a conditional mixture path guiding that explicitly computes optimal weights on the fly. Furthermore, we show how to perform positivization to eliminate sign variance and extend to scenes with millions of parameters.

To the best of our knowledge, this is the first generic framework for applying path guiding to differentiable rendering. Extensive experiments demonstrate that our method achieves nearly one order of magnitude improvements over state-of-the-art methods in terms of variance reduction in gradient estimation and errors of inverse optimization. The implementation of our proposed method is available at https://github.com/mollnn/conditional-mixture.

CCS Concepts: • **Computing methodologies** → **Rendering**.

ACM Reference Format:
Zhimin Fan, Pengcheng Shi, Mufan Guo, Ruoyu Fu, Yanwen Guo, and Jie Guo. 2024. Conditional Mixture Path Guiding for Differentiable Rendering. *ACM Trans. Graph.* 43, 4, Article 1 (August 2024), 11 pages. https://doi.org/10.1145/3618360

1 INTRODUCTION

Inverse rendering based on physically based differentiable rendering (PBDR) [Li 2022; Zhao et al. 2020] formulates the inversion process as an optimization loop which involves gradient-descent on a high-dimensional domain with millions of scene parameters [Jakob et al. 2022]. The differentiable renderer provides estimations of gradients while handling intricate interactions via stochastic sampling.

Even though modern optimizers are theoretically able to deal with noisy gradients, the quality of Monte Carlo estimations, including the variance and bias of both primal and adjoint renderings, decides the overall efficacy of the optimization. In the worst-case scenario, images with excessive noise will cause the optimization process to diverge [Nimier-David et al. 2022; Vicini et al. 2021; Zhang et al. 2023]. The problem becomes more severe with the fact that only a small amount of samples per pixel (typically 4-32) [Belhe et al. 2024; Nicolet et al. 2023] are used in each optimization step. Therefore, variance reduction is necessary to accelerate the inversion process and ensure convergence.

Recently, differential counterparts of numerous variance reduction techniques have been adapted into PBDR frameworks. For

*Corresponding authors.

Authors' addresses: Zhimin Fan, zhiminfan2002@gmail.com, State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China; Pengcheng Shi, unmei772@gmail.com, State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China; Mufan Guo, orlando128@163.com, State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China; Ruoyu Fu, fry@nju.edu.cn, State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China; Yanwen Guo, ywguo@nju.edu.cn, State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China; Jie Guo, guojie@nju.edu.cn, State Key Lab for Novel Software Technology, Nanjing University, Nanjing, China.

instance, the sampling of bidirectional scattering distribution function (BSDF) derivatives can be improved using antithetic sampling [Zeltner et al. 2021; Zhang et al. 2021a] and positivization [Belhe et al. 2024; Owen and Zhou 2000]. Special strategies have also been developed to handle geometric continuities, by using control variates [Bangaru et al. 2020; Loubet et al. 2019], path space guiding [Zhang et al. 2020], primary sample space guiding [Yan et al. 2022], or sample reuse [Chang et al. 2023; Wang et al. 2023]. Nevertheless, to our knowledge, the differential sampling strategy [Zeltner et al. 2021] of differential radiance for material parameters that completely handles global illumination remains a clear gap.

In this work, we propose a generic framework for applying path guiding to differentiable rendering. A crucial challenge is that the estimation of differential radiance requires an embedded estimation of ordinary radiance, which is now resolved using the same path [Vicini et al. 2021; Zhang et al. 2020]. To overcome this issue, we propose a mixture guiding scheme that employs two separated spatial-directional distributions to fit the primal and differential integrands. We show that the choice of the mixture weights has a great impact on the noise level of estimated derivatives. This inspires us to develop conditional mixture importance sampling, which uses adaptive mixture weights conditioned on the previously sampled vertices in the path. Since gradients can be real-valued vectors, we further employ positivization to eliminate sign variance and use L_1 norm to support multiple parameters.

Through these, we construct a practical pipeline that sufficiently exploits spatial and temporal coherence across simulated light path samples in inverse rendering. We evaluate our pipeline in a variety of challenging scenarios, including derivative estimation and inverse reconstruction. Both qualitative and quantitative comparisons demonstrate that our approach serves as a substantial advancement in noise reduction for differentiable rendering in terms of the variance of parameter gradients and the acceleration of inverse optimization.

In summary, our main contributions in this paper are:

- A mixture importance sampling method for estimating derivatives to material parameters, considering the same path is used for estimating differential and ordinary radiance.
- An optimal choice of conditional mixture weights for local directional sampling, combined with positivization to handle real-valued gradients.
- An efficient differentiable rendering pipeline that enables efficient material reconstructions in scenes with strong indirect illumination, caustics, and complex visibility.

2 RELATED WORKS

Physically based differentiable rendering. Recent interest in Monte Carlo differentiable rendering [Kato et al. 2020; Li 2022; Zhao et al. 2020] begins with [Li et al. 2018], which paves the way for handling global illumination [Zhang et al. 2019] within a general-proposed path tracing [Kajiya 1986; Veach 1997] framework.

Follow-up works propose efficient and accurate techniques for estimating the boundary integral through explicit edge sampling [Yan et al. 2022; Zhang et al. 2020], analytical visibility [Zhou et al. 2021], reparameterization [Loubet et al. 2019], and warped-area sampling [Bangaru et al. 2020]. Implicit geometric representations [Bangaru et al. 2022; Vicini et al. 2022] and volumetric radiative transfer [Nimier-David et al. 2022; Zhang et al. 2021b] are also effectively supported. Meanwhile, other works proposed practical algorithmic frameworks [Nimier-David et al. 2020] using constant memory and linear time [Vicini et al. 2021], as well as automatic differentiation systems and compilers [Jakob et al. 2022; Nimier-David et al. 2019]. All of these advancements have resulted in the widespread adoption of physically based differentiable rendering in numerous inverse rendering applications [Azinovic et al. 2019; Luan et al. 2021; Zhu et al. 2022] handling various light transport phenomena. Our method is largely agnostic to the underlying algorithm and application, which is compatible or orthogonal to most of the existing techniques and strategies [Bangaru et al. 2020; Belhe et al. 2024; Chang et al. 2023; Loubet et al. 2019; Nicolet et al. 2023; Vicini et al. 2022; Zeltner et al. 2021]. In our validations, we demonstrate higher-quality gradient estimations and faster optimization when applying our method to Path Replay Backpropagation (PRB) [Jakob et al. 2022; Vicini et al. 2021].

Sampling strategies for differentiable rendering. The efficiency of Monte Carlo differentiable rendering is highly dependent on the design of sampling strategies. Yet, sampling strategies designed for primal integrand may no longer work when sampling derivatives, necessitating the development of customized *differential sampling strategies* for variance reduction [Zeltner et al. 2021].

Early attempts applied antithetic sampling to differential BSDF sampling [Zeltner et al. 2021; Zhang et al. 2021a] leveraging the symmetry of derivatives. This is also claimed as a special case of positivization [Belhe et al. 2024; Owen and Zhou 2000]. However, these methods do not take the incident radiance fields into account.

In contrast to the analytical ones, another line of work benefits from historical samples. Recently, Chang et al. [2023] utilized positivized ReSTIR [Bitterli et al. 2020] to efficiently sample the differential radiance with respect to material parameters, which takes both the incident radiance and BSDF derivatives into account. Wang et al. [2023] applied ReSTIR to differentiable rendering for both material and geometric parameters. However, both of these methods only consider direct lighting. Concurrent to our work, Balint et al. [2023] propose a meta-estimator that combines proportional and finite-difference estimations to efficiently reuse past samples, which is largely agnostic to our importance sampling strategy.

Besides, specific variance reduction techniques have also been used in handling geometric discontinuities [Bangaru et al. 2020; Li et al. 2018; Loubet et al. 2019; Yan et al. 2022].

Our work is the first to concentrate on differential sampling strategy for material parameters under global illumination. To ensure robustness and adaptability, we leverage the strong capability of path guiding. This is especially advantageous in scenes involving substantial correlation, such as those with strong indirect illumination, glossy interreflection, and complex visibility.

Path guiding. Monte Carlo rendering techniques rely heavily on importance sampling when constructing light transport paths, and the most promising sampling distributions for forward rendering are obtained through learning scene priors. Existing methods use different distribution representations, such as Gaussian Mixture

Models (GMM) [Vorba et al. 2014] and quad-trees [Müller et al. 2017]. Some approaches further reduce the variance by considering the correlation between consecutive vertices [Dodik et al. 2022; Dong et al. 2023; Herholz et al. 2016; Müller et al. 2019; Ruppert et al. 2020; Schüßler et al. 2022] or directly employing high-dimensional structures [Guo et al. 2018; Reibold et al. 2018; Zheng and Zwicker 2019]. Besides, guiding samples towards the high-variance regions also benefits some challenging scenes [Rath et al. 2020].

For differentiable rendering, existing works only handle the guided sampling for estimating geometric derivatives in path space differentiable rendering [Yan et al. 2022; Zhang et al. 2020], while the material derivatives are largely under-explored. Developing a guiding method for derivatives with respect to material parameters is a non-trivial task. Specifically, since a single path is used for both ordinary and differential radiance estimation, new challenges arise in several aspects, both theoretical and algorithmic. Our work is the first attempt to fill this gap, proposing theoretical analysis along with a practical pipeline.

Importance sampling for gradient-based optimization. The use of importance sampling has become crucial in gradient-based optimization algorithms. Among different approaches, adaptive importance sampling methods periodically re-evaluate the importance during training and thereby often surpass the static ones. A widely adopted target is to minimize the expected squared L_2 norm of gradient estimations, which can be achieved by using a proposal distribution proportional to the L_2 norm of gradient [Hanchi and Stephens 2021; Stich et al. 2017; Zhu 2016].

Specifically designed for light transport simulation, path guiding follows the principle of adaptive importance sampling. Our method also uses the L_1 norm to compute the importance weight of each sample from their gradient estimation.

3 BACKGROUND

The intricate interplay between the convergence rate of inverse optimization techniques and the variance of Monte Carlo estimations has undergone extensive exploration within the realms of optimization and machine learning. While importance sampling, as an essential variance reduction technique, finds widespread utility in conventional forward rendering [Kondapaneni et al. 2019] and reinforcement learning [Llorente et al. 2021], its nuanced application to differentiable rendering necessitates special considerations.

Importance sampling. Consider the definite integral of a function f over some domain Ω and its N-sample Monte Carlo estimator $\langle I \rangle$:

$$\langle I \rangle = \left\langle \int_\Omega f(x)\,\mathrm{d}x \right\rangle = \frac{1}{N}\sum_{j=1}^{N}\frac{f(X_j)}{p(X_j)}, \tag{1}$$

where N denotes the number of samples, and $p(x)$ is the sampling probability density function (PDF). The variance of $\langle I \rangle$ heavily depends on the PDF $p(x)$. Conventionally, the variance tends to be small, when $p(x)$ is approximately proportional to the integrand $f(x)$. If $p(x) \propto f(x)$, the estimation will reach zero variance.

Differentiable rendering. Recent practical physically based differentiable rendering (PBDR) algorithms are mostly built on solving the recursive integral equation on the local directional space $\mathcal{S}^2$ at all shading points in the scene, resulting in the following differential rendering equation [Nimier-David et al. 2020; Vicini et al. 2021]:

$$\begin{aligned}\partial_\pi L_o(\boldsymbol{x},\boldsymbol{\omega}_o) &= \partial_\pi L_e(\boldsymbol{x},\boldsymbol{\omega}_o) + \\ &\int_{\mathcal{S}^2} \left(f_s(\boldsymbol{x},\boldsymbol{\omega}_i,\boldsymbol{\omega}_o)\partial_\pi L_i(\boldsymbol{x},\boldsymbol{\omega}_i) + \partial_\pi f_s(\boldsymbol{x},\boldsymbol{\omega}_i,\boldsymbol{\omega}_o) L_i(\boldsymbol{x},\boldsymbol{\omega}_i)\right)\mathrm{d}\boldsymbol{\omega}_i^{\perp},\end{aligned} \tag{2}$$

where π denotes a set of scene parameters (e.g., optical properties), and $\partial_\pi := \partial/\partial\pi$. Here, we initially assume π only involves material parameters. L_o, L_i, and L_e refer to the outgoing, incident, and emitted radiance at a shading point $\boldsymbol{x}$. f_s is the bidirectional scattering distribution function (BSDF), parameterized by the incident direction $\boldsymbol{\omega}_i$, the outgoing direction $\boldsymbol{\omega}_o$, and the position of the shading point $\boldsymbol{x}$. The above equation states that the target differential outgoing radiance $\partial_\pi L_o$ involves three parts:

(1) Differential radiance that is emitted from light sources whose primal emission depends on π,
(2) Differential radiance that scatters like ordinary radiance, conforming to the principles of the BSDF,
(3) Differential radiance that is also added on the shading point $\boldsymbol{x}$ whose BSDF depends on π.

Like that in conventional forward rendering, solving the above differential rendering equation in the context of Monte Carlo requires importance sampling of light paths to reduce the variance. For forward rendering, we generally guide the sampling according to incident radiance fields and BSDFs. However, for differentiable rendering, we must consider them and their differential counterparts simultaneously. Furthermore, the estimation of differential radiance requires an embedded estimation of ordinary radiance, which is now resolved using the same path [Vicini et al. 2021; Zhang et al. 2020]. This poses a new challenge for finding proper sampling PDFs to satisfy all integrals in PBDR. In this paper, we propose a general solution for importance sampling light paths for the differential rendering equation, focusing on the derivatives with respect to the material parameters.

4 IMPORTANCE SAMPLING PATH DERIVATIVES

We first theoretically analyze the importance sampling issues in the context of PBDR. Then, we design conditional mixture importance sampling that is able to satisfy all integrals in PBDR.

4.1 Motivation

To sample light paths for derivative estimation, prior works mostly use standard BSDF sampling [Vicini et al. 2021; Zhang et al. 2020], or design special tools for sampling the BSDF derivatives [Belhe et al. 2024; Zeltner et al. 2021; Zhang et al. 2021a], paying little attention to the incident radiance field, even when combined with direct light sampling strategies like next event estimation (NEE).

Recently, Chang et al. [2023] utilized positivized ReSTIR [Bitterli et al. 2020] to efficiently sample the differential emittance term, which for the first time takes both incident radiance fields and BSDF derivatives into account. However, they only handled direct lighting, due to the limitation of ReSTIR.

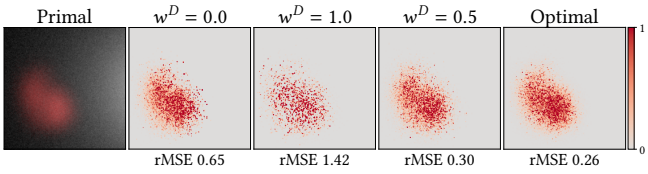

Fig. 2. Derivatives with respect to the bunny's reflectence under different weights in Eq. (5). Sampling paths using only p^L or p^D results in high variance. The relative mean squared error (rMSE) metric clearly shows the benefit of employing a mixture distribution of p^L and p^D.

In the context of forward rendering, to ensure low variance, estimating ordinary radiance requires sampling directions with probability density proportional to the integrand of the rendering equation, i.e.,

$$p^L(\boldsymbol{\omega}_i|\boldsymbol{x},\boldsymbol{\omega}_o) \propto f_s(\boldsymbol{x},\boldsymbol{\omega}_i,\boldsymbol{\omega}_o)L_i(\boldsymbol{x},\boldsymbol{\omega}_i). \tag{3}$$

However, the optimal sampling PDF for estimating the integral of the differential rendering equation is expected to satisfy:

$$p^D(\boldsymbol{\omega}_i|\boldsymbol{x},\boldsymbol{\omega}_o) \propto f_s(\boldsymbol{x},\boldsymbol{\omega}_i,\boldsymbol{\omega}_o)\partial_\pi L_i(\boldsymbol{x},\boldsymbol{\omega}_i)+\partial_\pi f_s(\boldsymbol{x},\boldsymbol{\omega}_i,\boldsymbol{\omega}_o)L_i(\boldsymbol{x},\boldsymbol{\omega}_i). \tag{4}$$

Modern PBDR algorithms use two separated passes for primal rendering and adjoint rendering to mitigate correlation and ensure unbiasedness [Nimier-David et al. 2020]. Thus, for primal renderings, light paths can be importance sampled using p^L defined in Eq. (3).

However, in the adjoint rendering pass, p^D does not always achieve the goal of importance sampling. This is because the estimation of differential radiance in Eq. (2) requires a recursive estimation of the ordinary radiance[1], and to significantly lower the cost in terms of memory and computation time, most recent PBDR methods share the same set of path samples when estimating the ordinary radiance and the differential radiance [Vicini et al. 2021; Zhang et al. 2020]. Since a path is used to estimate two different quantities, using only p^L or p^D to sample paths will cause extremely high variance in some cases for estimating the differential or ordinary radiance, resulting in severe noise. To address this issue, a straightforward solution is to use a mixture of the above two distributions.

4.2 Mixture sampling

For differentiable rendering, the mixture sampling PDF takes the form of

$$\begin{aligned} p^M(\boldsymbol{\omega}_i|\boldsymbol{x},\boldsymbol{\omega}_o) \propto\ & w^L f_s(\boldsymbol{x},\boldsymbol{\omega}_i,\boldsymbol{\omega}_o)L_i(\boldsymbol{x},\boldsymbol{\omega}_i)+ \\ & w^D\left(f_s(\boldsymbol{x},\boldsymbol{\omega}_i,\boldsymbol{\omega}_o)\partial_\pi L_i(\boldsymbol{x},\boldsymbol{\omega}_i) + \partial_\pi f_s(\boldsymbol{x},\boldsymbol{\omega}_i,\boldsymbol{\omega}_o)L_i(\boldsymbol{x},\boldsymbol{\omega}_i)\right). \end{aligned} \tag{5}$$

with w^L and w^D being the weights of the components. When setting the weights to fixed values (e.g., $w^L = w^D = 0.5$), we get one-sample Multiple Importance Sampling (MIS). Fig. 2 shows the benefits of using a mixture distribution on a typical scene involving strong indirect illumination. The comparisons show that such a simple mixture model can already reduce the variance of derivatives.

While pragmatic, this simple mixture sampling algorithm with fixed weights is suboptimal for differentiable rendering. Distinct vertices along diverse paths may require different optimal values of w^L and w^D, posing challenges in establishing a fixed value universally effective across all cases. For instance, a path connecting the camera, a vertex on a non-differentiable surface, and the light source sequentially may be frequently sampled by the simplified solution elucidated above. However, it has no contribution to the final estimations of the derivatives. An example scene demonstrating these phenomena is shown in Fig. 3.

[1] Here, the ordinary radiance estimation is only used to estimate the differential radiance, which is different from the ordinary radiance estimations in the primal rendering pass.

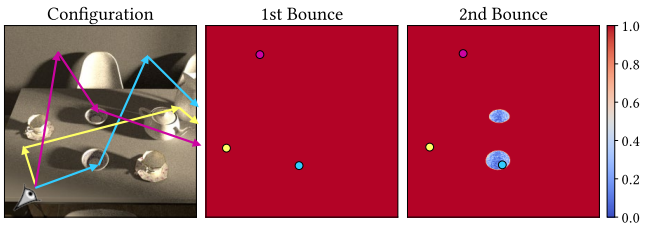

Fig. 3. Visualization of our optimal weights on the first and second bounce of each path. The derivatives are with respect to the reflectance of the two bowls. For the first bounce, we always sample p^D since $w_1^L = 0$, as sampling p^L could introduce path samples that never hit the bowl and have a zero contribution to gradients. The cyan path hits the target bowl on the first bounce, so when sampling the incident direction on the second bounce according to p^L, this path could have a non-zero gradient contribution, hence $w_2^L > 0$, $w_2^D > 0$. The magenta path and yellow path miss the bowl on the first bounce, thus $w_1^L = 0$, as sampling p^L could produce paths with zero gradient contribution (e.g., the yellow path).

From a path space perspective, importance sampling for differentiable rendering requires sampling each light path with probability density proportional to its contribution. For forward rendering, the contribution is a product of several factors. However, it no longer holds for differential rendering, where the contribution of a path is a finite sum of several terms, each of which is a product. This seems to require a path space or primary sample space guiding manner, which requires *high-dimensional* distributions and is problematic for long paths.

Our key observation is that the optimal proposal distribution can still be depicted as a mixture of various *local* directional distributions. Notably, the distribution is exclusively conditioned on the position of the shading point, while the mixture weights relies on the previously sampled path prefix. Since this new approach involves mixing several distributions with weights conditioned on prior vertices, we call it conditional mixture importance sampling.

4.3 Conditional mixture importance sampling

To derive the optimal mixture weights, we must take into account the entire path rather than focusing solely on the local sampling decision. To enhance clarity, we denote a path as $\overline{\boldsymbol{x}} = \boldsymbol{x}_1\boldsymbol{x}_2\ldots\boldsymbol{x}_n$. We further represent a path prefix as $\overleftarrow{\boldsymbol{x}_i} = \boldsymbol{x}_1\ldots\boldsymbol{x}_i$.

Given a path prefix $\overleftarrow{\boldsymbol{x}_i}$, our objective is to determine the target distribution for sampling the direction $\overrightarrow{\boldsymbol{x}_i\boldsymbol{x}_{i+1}}$. Note that this direction is employed in estimating both $L_o(\boldsymbol{x}_i, \overrightarrow{\boldsymbol{x}_i\boldsymbol{x}_{i-1}})$ and $\partial_\pi L_o(\boldsymbol{x}_i, \overrightarrow{\boldsymbol{x}_i\boldsymbol{x}_{i-1}})$. Consequently, we can express the contribution of the entire path

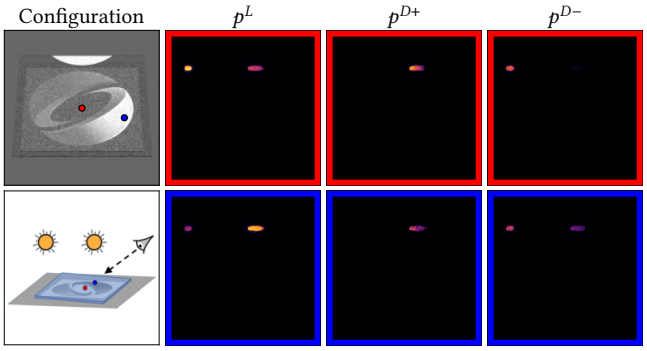

Fig. 4. Visualization of three directional distributions at two shading points, each fitting the primal integrand, positive gradients, and negative gradients, respectively. Derivatives are with respect to the roughness of the logo.

through these two components as

$$\partial_\pi \left(L_o \prod_{j=1}^{i-1} f_j \right) = L_o \, \partial_\pi \prod_{j=1}^{i-1} f_j + (\partial_\pi L_o) \prod_{j=1}^{i-1} f_j. \tag{6}$$

Here, L_o is the shorthand for $L_o(\boldsymbol{x}_i, \overrightarrow{\boldsymbol{x}_i \boldsymbol{x}_{i-1}})$ and f_i is the shorthand for the BSDF $f_s(\boldsymbol{x}_i, \overrightarrow{\boldsymbol{x}_i \boldsymbol{x}_{i+1}}, \overrightarrow{\boldsymbol{x}_i \boldsymbol{x}_{i-1}})$ of i-th vertex $\boldsymbol{x}_i$. Hence, the target distribution that includes both parts can be formulated as

$$\begin{aligned} &p(\boldsymbol{\omega}_i | \overleftarrow{\boldsymbol{x}_i}) \propto w^L(\overleftarrow{\boldsymbol{x}_i}) f_s(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) L_i(\boldsymbol{x}, \boldsymbol{\omega}_i) + \\ &w^D(\overleftarrow{\boldsymbol{x}_i}) \left(f_s(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) \partial_\pi L_i(\boldsymbol{x}, \boldsymbol{\omega}_i) + \partial_\pi f_s(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) L_i(\boldsymbol{x}, \boldsymbol{\omega}_i) \right). \end{aligned} \tag{7}$$

These two parts correspond to the two terms in Eq. (6). Thus, the weights are

$$w_i^L(\overleftarrow{\boldsymbol{x}_i}) = \partial_\pi \prod_{j=1}^{i-1} f_j, \quad w_i^D(\overleftarrow{\boldsymbol{x}_i}) = \prod_{j=1}^{i-1} f_j. \tag{8}$$

Sign variance elimination. The above analysis follows traditional definition of importance sampling that confined to positive scalar functions. In cases where the target is a real-valued scalar function, a common approach involves sampling with a probability proportional to the absolute value of the target distribution. However, this naive treatment tends to generate high variance as the samples possess importance weights with different signs. Consequently, real-valued functions cannot be perfectly importance sampled by a positive-valued PDF.

To resolve this issue, we apply the idea of positivization as follows. For a real-valued integrand $f(x)$, we construct two estimators for its positive and negative parts, respectively:

$$\left\langle \int_\Omega f(x) \mathrm{d}x \right\rangle = \frac{f^+(X^+)}{p^+(X^+)} + \frac{f^-(X^-)}{p^-(X^-)}. \tag{9}$$

Here, the samples X^+ and X^- are drawn from different proposal distributions $p^+(x)$ and $p^-(x)$, respectively.

A challenge is how to apply positivization to the recursive integral of the differential rendering equation. Applying it separately on each bounce will cause exponential branching, resulting in an exponential time complexity with respect to the length of the path.

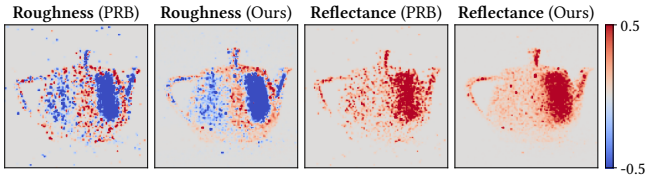

Fig. 5. Equal time (1.5 sec) comparison between PRB and our method, computing gradients to both the roughness and the reflectance. Our method utilizes the same distribution fitted from the L_1 norm of gradients.

We observe that the sign of differential radiance is determined only when it is emitted. When it is scattered, its sign will never change. This implies that we can decompose the contribution of a full path by decomposing only the emittance term, which is more easy to handle. Formally, we separate the adjoint rendering process into two passes. During the first pass, we clamp all the derivatives to non-negative values, resulting in a modified differential rendering equation:

$$\begin{aligned} \partial_\pi L_o^+(\boldsymbol{x}, \boldsymbol{\omega}_o) = &\int_{S^2} f_s(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) \partial_\pi L_i(\boldsymbol{x}, \boldsymbol{\omega}_i)^+ \, \mathrm{d}\boldsymbol{\omega}_i^\perp + \\ &\int_{S^2} \partial_\pi f_s^+(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) L_i(\boldsymbol{x}, \boldsymbol{\omega}_i) \, \mathrm{d}\boldsymbol{\omega}_i^\perp + \partial_\pi L_e^+(\boldsymbol{x}, \boldsymbol{\omega}_o). \end{aligned} \tag{10}$$

Similarly, during the second pass we clamp all the derivatives to non-positive values.

Then, we fit proposal distributions that involve derivatives separately for the two passes, i.e.,

$$p^{D+}(\boldsymbol{\omega}_i | \boldsymbol{x}, \boldsymbol{\omega}_o) \propto f_s(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) \partial_\pi L_i^+(\boldsymbol{x}, \boldsymbol{\omega}_i) + \partial_\pi f_s^+(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) L_i(\boldsymbol{x}, \boldsymbol{\omega}_i), \tag{11}$$

$$p^{D-}(\boldsymbol{\omega}_i | \boldsymbol{x}, \boldsymbol{\omega}_o) \propto f_s(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) \partial_\pi L_i^-(\boldsymbol{x}, \boldsymbol{\omega}_i) + \partial_\pi f_s^-(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) L_i(\boldsymbol{x}, \boldsymbol{\omega}_i). \tag{12}$$

An example of these distributions is shown in Fig. 4.

Additionally, we must consider the possibility of sign changes between consecutive iterations. Our solution is to mix a very small amount (5% in our implementation) of the negative distribution into the positive one, and vice versa.

Multiple parameters. The previous analysis assumes a single differentiable parameter π. The extension to multiple parameters is straightforward. Since now the gradient becomes a vector, we need to convert it into a scalar before fitting the sampling distributions. We currently use the L_1 norm of the gradient:

$$p^D(\boldsymbol{\omega}_i | \boldsymbol{x}, \boldsymbol{\omega}_o) \propto \| f_s(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) \partial_\pi L_i(\boldsymbol{x}, \boldsymbol{\omega}_i) + \partial_\pi f_s(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o) L_i(\boldsymbol{x}, \boldsymbol{\omega}_i) \|_1. \tag{13}$$

Special care should be taken when the parameters exhibit diverse units and thus the scales of their respective gradients may vary significantly. In this case, additional normalization steps are required. To ensure stability during inverse optimization, we normalize all parameter values to the interval $[0, 1]$ during the whole process.

Fig. 5 showcases an example that validates our solution. As observed, our method consistently produces low variance gradients, utilizing the same distribution to handle multiple parameters.

5 CONDITIONAL MIXTURE PATH GUIDING

Based on the above analysis and the proposed conditional mixture importance sampling algorithm, we develop a practical PBDR pipeline named conditional mixture path guiding. This new pipeline allows us to well handle many challenging cases (strong indirect illumination, caustics, and complex visibility) that are not solvable by previous works in inverse rendering.

In previous works, a separate training phase with hundreds of samples per pixel is used for path guiding in forward rendering, and the learned distribution is subsequently used during final rendering. However, for differentiable rendering, there are extremely few samples (i.e., ≤ 32) per pixel for each optimization step. This leads to many design decisions that deviate from the forward counterpart.

In our pipeline, there is only one guiding iteration for each optimization step. No separate training phase exists. Every sample is utilized for both fitting the guiding distributions and optimizing the scene parameters. Since it is impossible to fit distributions independently for every optimization step, the distributions are updated incrementally over all the iterations.

Sample gathering. We use the path samples gathered in the previous rendering iteration to update the guiding distribution. We follow the general pipeline of PRB [Vicini et al. 2021] in which each iteration consists of three rendering passes. The first is a primal rendering pass like conventional forward rendering for the loss computation. While it is still primitive, the second pass gathers data for the third pass, which backpropagates adjoint radiance throughout the scene using the same set of samples as in the second pass. We use the samples from the first and the second pass to fit p^L, and the last pass to fit p^D.

Distribution fitting. We use KD-tree as the spatial structure and adopt the adaptive quadtree to represent directional distributions, with stochastic spatial filtering and box directional filtering enabled [Müller 2019]. We subdivide a spatial node if the number of primal samples inside a spatial cell is greater than 32000. A directional node will be subdivided only if it contains more than 1% flux of the whole quadtree [Müller et al. 2017]. Each spatial node contains three quadtrees, corresponding to p^L, p^{D+}, and p^{D-} respectively.

We update all the structures and distributions upon the completion of an entire optimization step. We sequentially refine the tree structures, splat the samples into the tree, and then build the distribution by setting the value of non-leaf nodes to the sum of its children. The path samples in each optimization step are generated using a copy of the distribution before updating [Müller et al. 2017]. For inverse rendering, the scene is changing with every iteration. To fit an accurate sampling distribution for the current iteration, the samples gathered from previous iterations are usually less useful. For this reason, we use a constant decay weight (i.e., 0.9) in an exponential weighting scheme for distributions over iterations.

Importance sampling. To determine the sampling directions, we first use Eq. (8) in Sec. 4.3 to compute the mixture weights. Then, we randomly choose one of the two distributions according to the weights and draw samples from the chosen distribution. It is necessary to query the two distributions to obtain a weighted average before evaluating the probability density. Note that during the primal rendering phase, we sample paths simply using p^L. Besides, we always undertake a 25% defensive sampling (with regular BSDF sampling in our implementation) to avoid an arbitrarily small probability density and keep the estimator unbiased for both primal and adjoint rendering passes, following the path guiding convention [Vorba et al. 2019]. For the first iteration, without guiding distributions available, we exclusively perform BSDF sampling.

Table 1. Experimental setup. We show the average time of kernel execution for each iteration using Mitsuba's megakernel mode. $n + m$ SPP means using n spp for primal rendering and m app for adjoint rendering.

Figure	PRB		Ours		
	SPP	Time	SPP	Time (render)	Time (fit)
Fig. 1	16 + 16	1.878	16 + 16	2.271	0.687
Fig. 9	24 + 4	1.797	16 + 4	1.928	0.815
Fig. 12	40 + 2	0.338	16 + 2	0.308	0.165

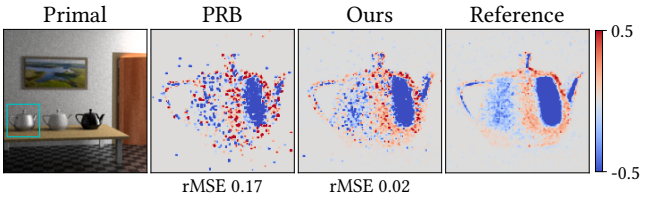

Fig. 6. Equal sample comparison (16 spp) comparing our derivative estimation with respect to the teapot's roughness with PRB at the 10th iteration.

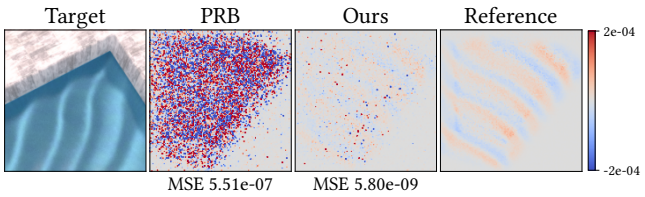

Fig. 7. Equal time comparison (2 sec) between our method and PRB. We show the gradients with respect to the normals backpropagated to textures. The guiding distribution is fitted through 10 iterations. The roughness of the water surface is 0.1.

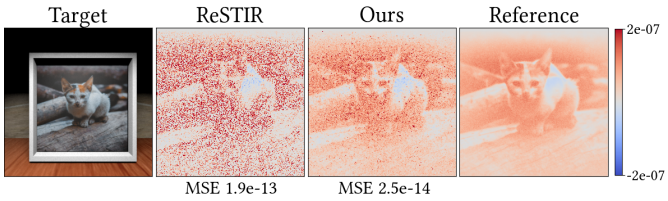

Fig. 8. Comparison with parameter-space ReSTIR [Chang et al. 2023]. We compute derivatives to the base color of the paint. The scene configuration is shown in Fig. 10.

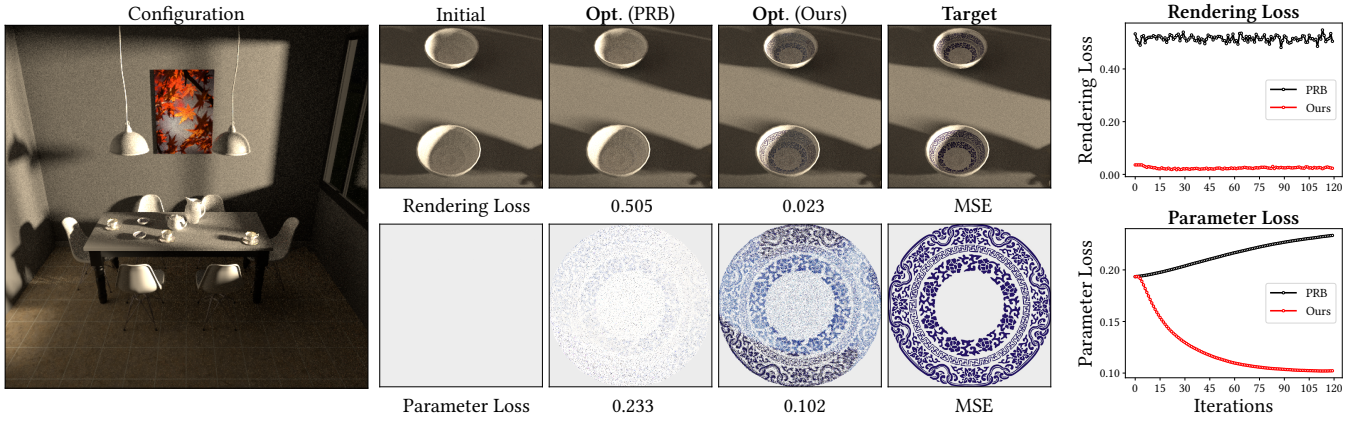

Fig. 9. Optimizing spatially varying base color for two bowls using a single view. The scene includes strong indirect illumination.

6 RESULTS

We have implemented our method in Mitsuba 3 [Jakob et al. 2022]. The rendering parts run on the CPU with the LLVM backend, while the distribution fitting part uses a standalone C++ implementation. Distributions are stored as 1D textures in Dr.Jit. Each tree node corresponds to a texel storing values and indices of its children. The indices of texels correspond to the indices of tree nodes. When sampling and querying distributions, the texture is queried at each level of the tree to select one of the child nodes and compute the probability density in the corresponding region. We pad the texture size to a fixed value to avoid mega-kernel recompilation. For gradient estimation, we show the relative mean squared error (rMSE) with a ground truth rendered using PRB at high sample rates. For inverse rendering comparison, we show rendering loss and parameter loss. We use MSE as the loss function and Adam optimizer without momentum. All timings are conducted on a PC with Intel i9-13900KF 24-core CPU and NVIDIA RTX 4080 GPU.

We evaluate our method in scenes with complex light transport effects, such as high-frequency caustics, strong indirect illumination, and complex visibility. The experimental setting is shown in Table 1. We plot the loss except for the first four iterations to circumvent extremely high values since spatial structures are not yet subdivided thoroughly [Müller 2019]. One can alternatively treat the first few iterations as a burn-in phase [Wang et al. 2023].

Gradient estimation. In Fig. 6, we compare our method with PRB using BSDF sampling with Next-Event Estimation (NEE). The scene is a well-known example characterized by strong indirect illumination. Our evaluation involves gradient estimation in the forward mode of the teapot's roughness. Here, a path with a non-zero gradient contribution must connect the teapot and the light source positioned behind the door. Detached BSDF sampling and NEE of PRB struggle to sample such paths, resulting in noisy gradient estimation. Our method can fit the distribution from historical samples and guide the sampling towards high-contribution regions. As a result, our method achieves significantly lower variance estimation than PRB.

Texture optimization. The convergence of optimization using different estimators in a simple optimization task is shown in Fig. 9. We use a single view to optimize the spatially variable base color of two bowls. The scene contains strong indirect illumination: the sunlight passing through the window is occluded, so the bowls only receive indirect lighting. We compare inverse optimization using two estimators: PRB using detached BSDF sampling with NEE and our mixture guiding estimator. In such a setting, the BSDF sampling and NEE of PRB fail to reach paths that connect the diffuse shading point inside the bowl to the light source passing through some diffuse reflector and the glass window. Consequently, PRB generates extremely sparse derivatives, which makes the optimization process diverge. Thanks to the strong capability of path guiding in handling indirect illumination, our method significantly reduces the variance of gradient estimation as shown in Fig. 7, resulting in stable convergence and clear details in the reconstructed textures.

Caustics optimization. Fig. 1 showcases the reconstruction of the normal map texture of the water in a pool. This scene involves high-frequency underwater caustics caused by waves on the near-specular surface. The underlying simulation involves the well-known SDS paths, which is known as a failure case for unidirectional or bidirectional path tracing. In such a setting, the BSDF sampling and NEE of PRB fail to generate paths with high contribution. Our approach can learn and importance sample high-frequency transport leading to reduced variance in both primal rendering and gradient estimation. As a result, our method is able to recover the caustic patterns close to the references.

Comparison with parameter-space ReSTIR. In Fig. 10, we compare our method with parameter-space ReSTIR [Chang et al. 2023]. We optimize the base color of the paint receiving indirect lighting passing through a hole and show the derivatives in Fig. 8. Since parameter-space ReSTIR only implements the sample reuse for the adjoint rendering phase, we use high sample rates for both two methods, such that the quality difference in primal rendering is small. Even in this setting, parameter-space ReSTIR still produces much more noise in the gradients than ours, resulting in slower

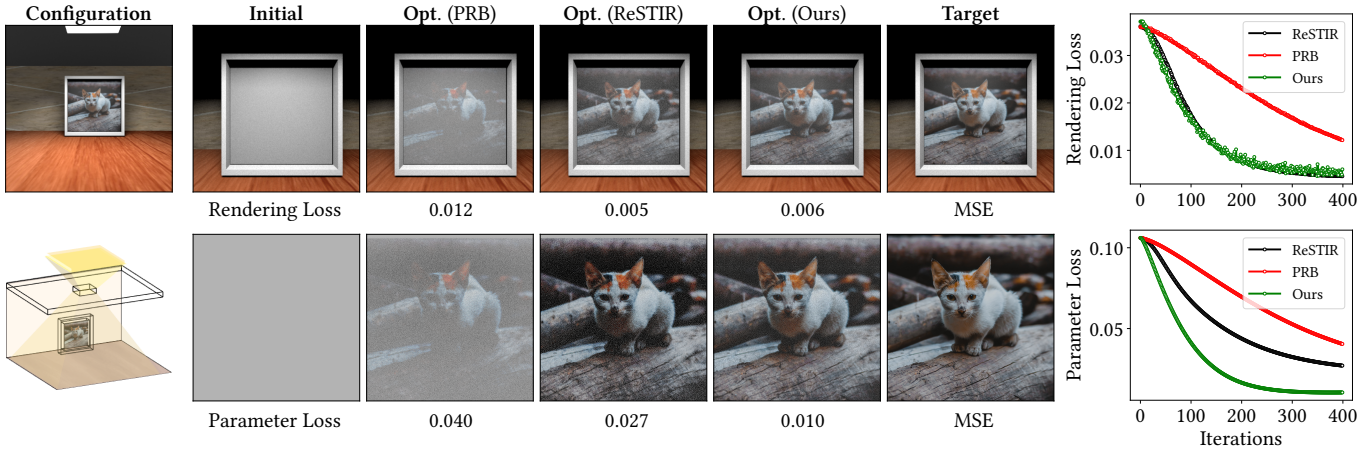

Fig. 10. Comparing with parameter-space ReSTIR in reconstructing the paint's base color under direct lighting passing through a hole.

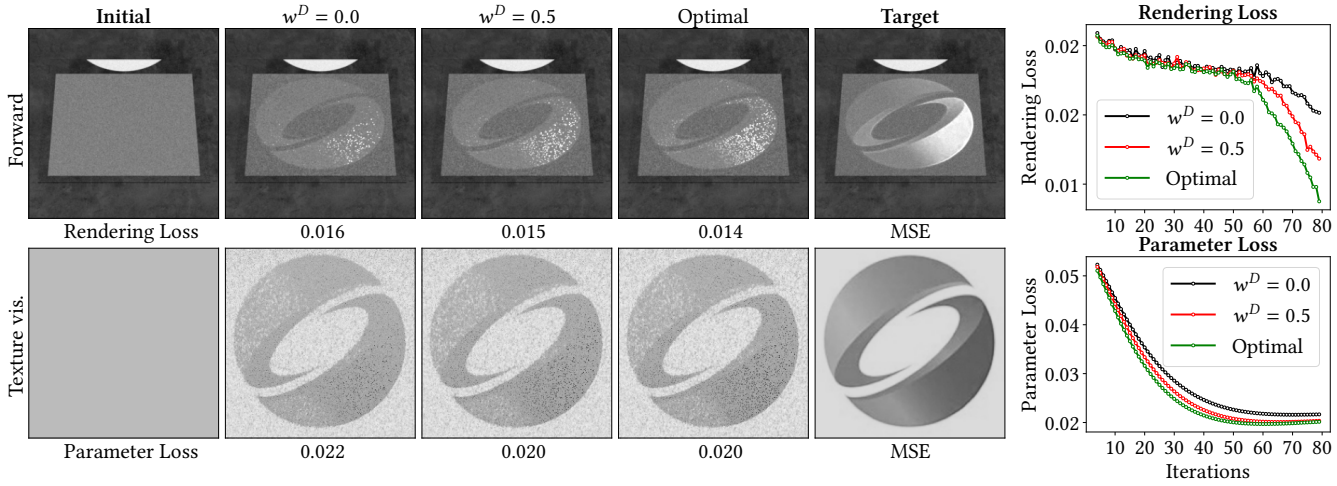

Fig. 11. Comparing various mixture sampling methods by optimizing the roughness texture of the rough conductor plane to reconstruct the SIGGRAPH logo.

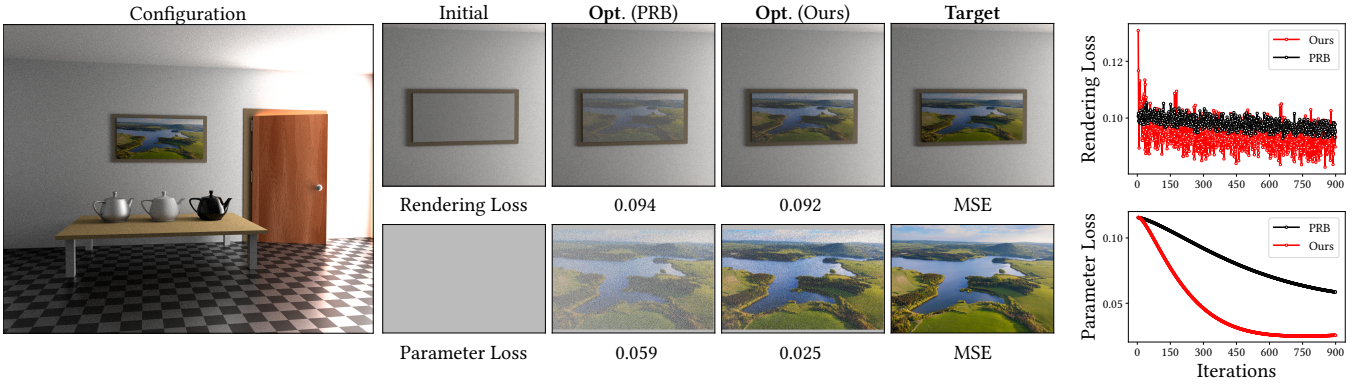

Fig. 12. In Veach Ajar, a well-known scene with prominent indirect illumination, we reconstruct the base color of the painting on the wall.

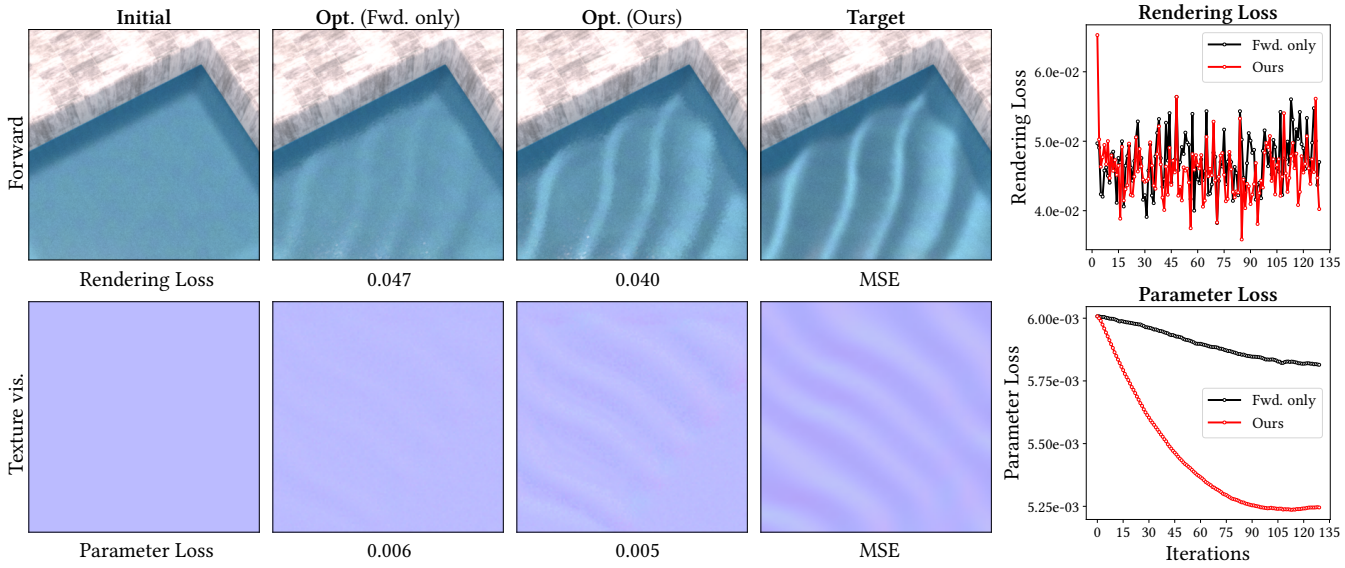

Fig. 13. We compare our complete method with a variant that utilizes path guiding solely to enhance rendering quality in the primal phase, denoted as "Fwd. only". We optimize the normal map of the water surface whose roughness is 0.1.

optimization convergence. This disparity arises from the fact that in that approach, sample reuse is performed independently in each parameter space. In contrast, our guiding distribution is shared within each leaf node of the spatial structure, leveraging spatial coherence more effectively.

Validation of mixture sampling. In Fig. 2, we show the derivatives rendered using two independent distributions and their mixture. Using p^L only causes a high variance for estimating the differential radiance since the differential quantities are not considered. On the other hand, using p^D also results in very noisy estimation. This is because estimating the derivatives also requires estimating the ordinary radiance, but the differential integrand may be extremely small in some regions where the primal integrand is large (e.g., Fig. 4). A fixed mixture sampling achieves a proper balance between these two parts, resulting in significantly lower variance.

Validation of optimal mixture weights. We validate our proposed mixture importance sampling against the fixed mixture in Fig. 11. Our conditional mixture path guiding successfully predicts the optimal weight and produces derivative estimation with much less noise than various constant weighting schemes. Consequently, our method leads to faster convergence of inverse optimization.

Validation of positivization. We validate our application of positivization in Fig. 14. When our guiding pipeline is employed without positivization, the target function is set as the absolute value of the integrand. The scene incorporates two light sources from different angles, introducing both positive and negative gradients, respectively. As seen, the variant without positivization results in a higher variance in gradient estimate due to sign changes. Our application

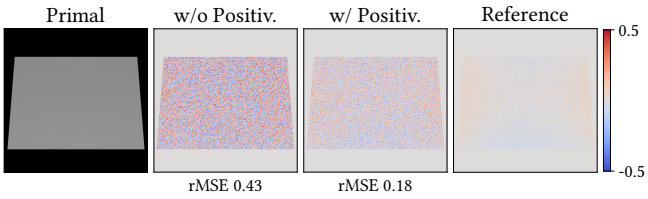

Fig. 14. Positivization significantly reduces sign variance, generating gradients closer to the reference. The derivative is with respect to the roughness of the conductor plane.

of positivization significantly reduces the sign variance, producing gradients more akin to the reference with negligible overhead.

Forward only vs. our complete model. In the context of inverse rendering, one can utilize path guiding solely to enhance the rendering quality in the primal phase, while keeping the sampling in the adjoint phase unchanged. We compare this variant, denoted as "Fwd. only" with our complete method in Fig. 13. Guiding the primal rendering reduces the variance in the primal rendering, leading to decreased variance and bias in the derivative of pixel intensity with respect to radiance. However, the variance in the derivative of radiance with respect to scene parameters remains high. In contrast, our complete method improves both parts and thus facilitates smooth convergence during optimization, resulting in clearer reconstructed caustic patterns.

Additionally, Fig. 15 compares our approach with a naïve application of path guiding using solely p^L with positivization. Our method demonstrates visible variance reduction compared to this baseline, which validates the importance of conditional mixture sampling.

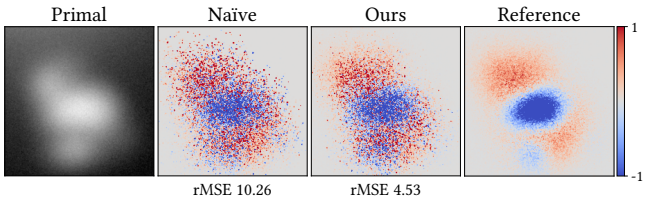

Fig. 15. Conditional mixture sampling achieves more variance reduction than a direct application of conventional path guiding and positivization. The derivative is with respect to the roughness of the bunny.

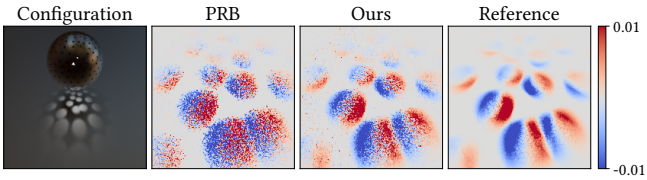

Fig. 16. Estimation of derivatives with respect to the ball's rotation angle by combining warped-area sampling [Bangaru et al. 2020].

Geometric derivatives. Although our initial focus is material derivatives, our method is also beneficial for geometric derivatives. We apply our method to reparameterized PRB, where warped-area sampling [Bangaru et al. 2020] is employed to estimate boundary terms. The only difference is that the BSDF derivatives $\partial_\pi f_s(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$ is replaced by $f_s(\boldsymbol{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o)\partial_\pi \|J_R(\boldsymbol{x}, \pi)\|$, where R is the reparameterization and J_R is its Jacobian [Zeltner et al. 2021]. As demonstrated in Fig. 16, our method can still guide samples towards high-contribution regions, leading to significant variance reduction.

7 DISCUSSIONS

Hyper-parameters and design choices. Our guiding pipeline is generic and largely extendable with many design choices and hyper-parameters. For instance, setting a low spatial subdivision threshold could lead to noisy directional distributions due to insufficient samples available for fitting. Conversely, a high threshold may result in coarse distributions and inefficiency, especially in scenarios with strong spatial-directional correlations (e.g., Fig. 1).

Our method introduces one extra parameter, the decay rate of distributions across optimization steps. The choice should be informed by the magnitude of change between iterations. The optimal choice could also be inspired by recent works [Nicolet et al. 2023].

Difference between consecutive iterations. Since the distribution is fitted using samples from the previous few iterations, how to confront a significant change in between iterations still needs to be explored [Balint et al. 2023; Chang et al. 2023].

Failure cases. Guided sampling is not guaranteed to perform better than BSDF sampling, especially when the incident radiance field is low-frequency. Besides, in some cases like caustics cast by specular surfaces and tiny area light, the radiance distributions cannot be fitted well. An example is shown in Fig. 17, where the roughness of the water surface is much lower than that of Fig. 1. How to cope with such situations is an interesting topic for further study.

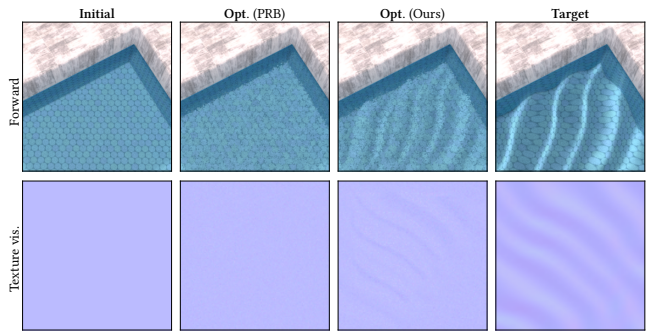

Fig. 17. A failure case of our method due to high-frequency radiance fields cast by near-specular surfaces and tiny area light source. We optimize the normal map of the water surface, whose roughness is 0.01.

8 CONCLUSION

Our work presents the first guided importance sampling technique for differentiable rendering focused on material parameters. Simply adopting path guiding for forward rendering may perform poorly since differentiable rendering involves estimating the ordinary and differential radiance simultaneously and usually utilizes the same path for both of them. In light of this, we propose a mixture sampling strategy and determine the optimal weight, which is conditioned on the vertices that have already been sampled. We further leverage positivization to achieve theoretical zero-variance convergence and support multiple parameters using the L_1 norm of gradients. We develop an effective pipeline that employs samples from successive gradient descent iterations to fit distributions incrementally. This serves as a novel adaptation of the path guiding to differentiable rendering, which leads to significantly lower variance and better convergence. We believe our work signifies a foundational step in devising sampling strategies for material derivatives under global illumination, suggesting promising avenues for variance reduction in the realms of differentiable and inverse rendering.

ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their suggestions. We also thank Pengpei Hong for the invaluable discussions. This work was supported by the National Natural Science Foundation of China (No. 61972194 and No. 62032011).

REFERENCES

Dejan Azinovic, Tzu-Mao Li, Anton Kaplanyan, and Matthias Niessner. 2019. Inverse Path Tracing for Joint Material and Lighting Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Martin Balint, Karol Myszkowski, Hans-Peter Seidel, and Gurprit Singh. 2023. Joint Sampling and Optimisation for Inverse Rendering. In *SIGGRAPH Asia 2023 Conference Papers (SA '23)*. Association for Computing Machinery, New York, NY, USA, Article 29, 10 pages. https://doi.org/10.1145/3610548.3618244

Sai Praveen Bangaru, Michael Gharbi, Fujun Luan, Tzu-Mao Li, Kalyan Sunkavalli, Milos Hasan, Sai Bi, Zexiang Xu, Gilbert Bernstein, and Fredo Durand. 2022. Differentiable Rendering of Neural SDFs through Reparameterization. In *SIGGRAPH Asia 2022 Conference Papers (SA '22)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3550469.3555397

Sai Praveen Bangaru, Tzu-Mao Li, and Frédo Durand. 2020. Unbiased Warped-Area Sampling for Differentiable Rendering. *ACM Transactions on Graphics* 39, 6 (Nov. 2020), 245:1–245:18. https://doi.org/10.1145/3414685.3417833

Yash Belhe, Bing Xu, Sai Praveen Bangaru, Ravi Ramamoorthi, and Tzu-Mao Li. 2024. Importance Sampling BRDF Derivatives. *ACM Trans. Graph.* (feb 2024). https://doi.org/10.1145/3648611 Just Accepted.

Benedikt Bitterli, Chris Wyman, Matt Pharr, Peter Shirley, Aaron Lefohn, and Wojciech Jarosz. 2020. Spatiotemporal Reservoir Resampling for Real-Time Ray Tracing with Dynamic Direct Lighting. *ACM Transactions on Graphics* 39, 4 (Aug. 2020). https://doi.org/10.1145/3386569.3392481

Wesley Chang, Venkataram Sivaram, Derek Nowrouzezahrai, Toshiya Hachisuka, Ravi Ramamoorthi, and Tzu-Mao Li. 2023. Parameter-space ReSTIR for Differentiable and Inverse Rendering. In *ACM SIGGRAPH 2023 Conference Proceedings (SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 18, 10 pages. https://doi.org/10.1145/3588432.3591512

Ana Dodik, Marios Papas, Cengiz Öztireli, and Thomas Müller. 2022. Path Guiding Using Spatio-Directional Mixture Models. *Computer Graphics Forum* 41, 1 (Feb. 2022), 172–189. https://doi.org/10.1111/cgf.14428

Honghao Dong, Guoping Wang, and Sheng Li. 2023. Neural Parametric Mixtures for Path Guiding. In *ACM SIGGRAPH 2023 Conference Proceedings (SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 29, 10 pages. https://doi.org/10.1145/3588432.3591533

Jerry Jinfeng Guo, Pablo Bauszat, Jacco Bikker, and Elmar Eisemann. 2018. Primary sample space path guiding. In *Proceedings of the Eurographics Symposium on Rendering: Experimental Ideas & Implementations* (Karlsruhe, Germany) *(SR '18)*. Eurographics Association, Goslar, DEU, 73–82. https://doi.org/10.2312/sre.20181174

Ayoub El Hanchi and David A. Stephens. 2021. Stochastic Reweighted Gradient Descent. In *International Conference on Machine Learning*. https://api.semanticscholar.org/CorpusID:232320705

Sebastian Herholz, Oskar Elek, Jiří Vorba, Hendrik Lensch, and Jaroslav Křivánek. 2016. Product Importance Sampling for Light Transport Path Guiding. *Computer Graphics Forum* 35, 4 (2016), 67–77. https://doi.org/10.1111/cgf.12950

Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, and Delio Vicini. 2022. DR.JIT: A Just-in-Time Compiler for Differentiable Rendering. *ACM Transactions on Graphics* 41, 4 (July 2022), 124:1–124:19. https://doi.org/10.1145/3528223.3530099

James T. Kajiya. 1986. The Rendering Equation. *Proceedings of the 13th annual conference on Computer graphics and interactive techniques - SIGGRAPH '86* (1986), 143–150. https://doi.org/10.1145/15922.15902

Hiroharu Kato, Deniz Beker, Mihai Morariu, Takahiro Ando, Toru Matsuoka, Wadim Kehl, and Adrien Gaidon. 2020. Differentiable Rendering: A Survey. https://doi.org/10.48550/arXiv.2006.12057 arXiv:2006.12057 [cs]

Ivo Kondapaneni, Petr Vevoda, Pascal Grittmann, Tomáš Skřivan, Philipp Slusallek, and Jaroslav Křivánek. 2019. Optimal Multiple Importance Sampling. *ACM Transactions on Graphics* 38, 4 (Aug. 2019), 1–14. https://doi.org/10.1145/3306346.3323009

Tzu-Mao Li. 2022. Differentiable Visual Computing: Challenges and Opportunities. *IEEE Computer Graphics and Applications* 42, 2 (March 2022), 101–109. https://doi.org/10.1109/MCG.2022.3149550

Tzu-Mao Li, Miika Aittala, Frédo Durand, and Jaakko Lehtinen. 2018. Differentiable Monte Carlo Ray Tracing through Edge Sampling. *ACM Transactions on Graphics* 37, 6 (Dec. 2018), 222:1–222:11. https://doi.org/10.1145/3272127.3275109

F. Llorente, L. Martino, J. Read, and D. Delgado. 2021. A Survey of Monte Carlo Methods for Noisy and Costly Densities with Application to Reinforcement Learning. https://doi.org/10.48550/arXiv.2108.00490 arXiv:2108.00490 [cs, stat]

Guillaume Loubet, Nicolas Holzschuch, and Wenzel Jakob. 2019. Reparameterizing Discontinuous Integrands for Differentiable Rendering. *ACM Transactions on Graphics* 38, 6 (Nov. 2019), 228:1–228:14. https://doi.org/10.1145/3355089.3356510

Fujun Luan, Shuang Zhao, Kavita Bala, and Zhao Dong. 2021. Unified Shape and SVBRDF Recovery Using Differentiable Monte Carlo Rendering. *Computer Graphics Forum* 40, 4 (2021), 101–113. https://doi.org/10.1111/cgf.14344

Thomas Müller. 2019. "Practical Path Guiding" in Production. In *ACM SIGGRAPH Courses: Path Guiding in Production, Chapter 10* (Los Angeles, California). ACM, New York, NY, USA, 18:35–18:48. https://doi.org/10.1145/3305366.3328091

Thomas Müller, Markus Gross, and Jan Novák. 2017. Practical Path Guiding for Efficient Light-Transport Simulation. *Computer Graphics Forum* 36, 4 (July 2017), 91–100. https://doi.org/10.1111/cgf.13227

Thomas Müller, Brian Mcwilliams, Fabrice Rousselle, Markus Gross, and Jan Novák. 2019. Neural Importance Sampling. *ACM Transactions on Graphics* 38, 5 (Oct. 2019), 1–19. https://doi.org/10.1145/3341156

Baptiste Nicolet, Fabrice Rousselle, Jan Novak, Alexander Keller, Wenzel Jakob, and Thomas Müller. 2023. Recursive Control Variates for Inverse Rendering. *ACM Trans. Graph.* 42, 4, Article 62 (jul 2023), 13 pages. https://doi.org/10.1145/3592139

Merlin Nimier-David, Thomas Müller, Alexander Keller, and Wenzel Jakob. 2022. Unbiased Inverse Volume Rendering with Differential Trackers. *ACM Transactions on Graphics* 41, 4 (July 2022), 1–20. https://doi.org/10.1145/3528223.3530073

Merlin Nimier-David, Sébastien Speierer, Benoît Ruiz, and Wenzel Jakob. 2020. Radiative Backpropagation: An Adjoint Method for Lightning-Fast Differentiable Rendering. *ACM Transactions on Graphics* 39, 4 (Aug. 2020), 146:146:1–146:146:15. https://doi.org/10.1145/3386569.3392406

Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. 2019. Mitsuba 2: A Retargetable Forward and Inverse Renderer. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 38, 6 (Dec. 2019). https://doi.org/10.1145/3355089.3356498

Art Owen and Yi Zhou. 2000. Safe and Effective Importance Sampling. *J. Amer. Statist. Assoc.* 95, 449 (March 2000), 135–143. https://doi.org/10.1080/01621459.2000.10473909

Alexander Rath, Pascal Grittmann, Sebastian Herholz, Petr Vévoda, Philipp Slusallek, and Jaroslav Křivánek. 2020. Variance-Aware Path Guiding. *ACM Transactions on Graphics* 39, 4 (Aug. 2020). https://doi.org/10.1145/3386569.3392441

Florian Reibold, Johannes Hanika, Alisa Jung, and Carsten Dachsbacher. 2018. Selective Guided Sampling with Complete Light Transport Paths. *ACM Transactions on Graphics* 37, 6 (Dec. 2018), 1–14. https://doi.org/10.1145/3272127.3275030

Lukas Ruppert, Sebastian Herholz, and Hendrik P. A. Lensch. 2020. Robust Fitting of Parallax-Aware Mixtures for Path Guiding. *ACM Transactions on Graphics* 39, 4 (Aug. 2020). https://doi.org/10.1145/3386569.3392421

Vincent Schüßler, Johannes Hanika, Alisa Jung, and Carsten Dachsbacher. 2022. Path Guiding with Vertex Triplet Distributions. *Computer Graphics Forum* 41, 4 (July 2022), 1–15. https://doi.org/10.1111/cgf.14582

Sebastian U. Stich, Anant Raj, and Martin Jaggi. 2017. Safe Adaptive Importance Sampling. https://doi.org/10.48550/arXiv.1711.02637 arXiv:1711.02637 [cs, math]

Eric Veach. 1997. Robust Monte Carlo Methods for Light Transport Simulation.

Delio Vicini, Sébastien Speierer, and Wenzel Jakob. 2021. Path Replay Backpropagation: Differentiating Light Paths Using Constant Memory and Linear Time. *ACM Transactions on Graphics* 40, 4 (Aug. 2021), 1–14. https://doi.org/10.1145/3450626.3459804

Delio Vicini, Sébastien Speierer, and Wenzel Jakob. 2022. Differentiable Signed Distance Function Rendering. *ACM Transactions on Graphics* 41, 4 (July 2022), 125:1–125:18. https://doi.org/10.1145/3528223.3530139

Jiří Vorba, Johannes Hanika, Sebastian Herholz, Thomas Müller, Jaroslav Křivánek, and Alexander Keller. 2019. Path Guiding in Production. In *ACM SIGGRAPH 2019 Courses*. ACM, Los Angeles California, 1–77. https://doi.org/10.1145/3305366.3328091

Jiří Vorba, Ondřej Karlík, Martin Šik, Tobias Ritschel, and Jaroslav Křivánek. 2014. On-Line Learning of Parametric Mixture Models for Light Transport Simulation. *ACM Transactions on Graphics* 33, 4 (July 2014), 1–11. https://doi.org/10.1145/2601097.2601203

Yu-Chen Wang, Chris Wyman, Lifan Wu, and Shuang Zhao. 2023. Amortizing Samples in Physics-Based Inverse Rendering Using ReSTIR. *ACM Trans. Graph.* 42, 6, Article 214 (dec 2023), 17 pages. https://doi.org/10.1145/3618331

Kai Yan, Christoph Lassner, Brian Budge, Zhao Dong, and Shuang Zhao. 2022. Efficient Estimation of Boundary Integrals for Path-Space Differentiable Rendering. *ACM Transactions on Graphics* 41, 4 (July 2022), 123:1–123:13. https://doi.org/10.1145/3528223.3530080

Tizian Zeltner, Sébastien Speierer, Iliyan Georgiev, and Wenzel Jakob. 2021. Monte Carlo estimators for differential light transport. *ACM Trans. Graph.* 40, 4, Article 78 (jul 2021), 16 pages. https://doi.org/10.1145/3450626.3459807

Cheng Zhang, Zhao Dong, Michael Doggett, and Shuang Zhao. 2021a. Antithetic Sampling for Monte Carlo Differentiable Rendering. *ACM Transactions on Graphics* 40, 4 (July 2021), 77:1–77:12. https://doi.org/10.1145/3450626.3459783

Cheng Zhang, Bailey Miller, Kai Yan, Ioannis Gkioulekas, and Shuang Zhao. 2020. Path-Space Differentiable Rendering. *ACM Transactions on Graphics* 39, 4 (Aug. 2020), 143:143:1–143:143:19. https://doi.org/10.1145/3386569.3392383

Cheng Zhang, Lifan Wu, Changxi Zheng, Ioannis Gkioulekas, Ravi Ramamoorthi, and Shuang Zhao. 2019. A Differential Theory of Radiative Transfer. *ACM Transactions on Graphics* 38, 6 (Nov. 2019), 227:1–227:16. https://doi.org/10.1145/3355089.3356522

Cheng Zhang, Zihan Yu, and Shuang Zhao. 2021b. Path-Space Differentiable Rendering of Participating Media. *ACM Transactions on Graphics* 40, 4 (July 2021), 76:1–76:15. https://doi.org/10.1145/3450626.3459782

Ziyi Zhang, Nicolas Roussel, and Wenzel Jakob. 2023. Projective Sampling for Differentiable Rendering of Geometry. *ACM Trans. Graph.* 42, 6, Article 212 (dec 2023), 14 pages. https://doi.org/10.1145/3618385

Shuang Zhao, Wenzel Jakob, and Tzu-Mao Li. 2020. Physics-Based Differentiable Rendering: From Theory to Implementation. In *ACM SIGGRAPH 2020 Courses*. ACM, Virtual Event USA, 1–30. https://doi.org/10.1145/3388769.3407454

Quan Zheng and Matthias Zwicker. 2019. Learning to Importance Sample in Primary Sample Space. *Computer Graphics Forum* 38, 2 (May 2019), 169–179. https://doi.org/10.1111/cgf.13628

Yang Zhou, Lifan Wu, Ravi Ramamoorthi, and Ling-Qi Yan. 2021. Vectorization for Fast, Analytic, and Differentiable Visibility. *ACM Transactions on Graphics* 40, 3 (July 2021), 27:1–27:21. https://doi.org/10.1145/3452097

Jingsen Zhu, Fujun Luan, Yuchi Huo, Zihao Lin, Zhihua Zhong, Dianbing Xi, Rui Wang, Hujun Bao, Jiaxiang Zheng, and Rui Tang. 2022. Learning-Based Inverse Rendering of Complex Indoor Scenes with Differentiable Monte Carlo Raytracing. In *SIGGRAPH Asia 2022 Conference Papers (SA '22)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3550469.3555407

Rong Zhu. 2016. Gradient-Based Sampling: An Adaptive Importance Sampling for Least-squares. https://doi.org/10.48550/arXiv.1803.00841 arXiv:1803.00841 [cs, stat]