Particle Rendering: Implicitly Aggregating Incident and Outgoing Light Fields for Novel View Synthesis

Tao Hu [1] Zhiwen Yan [1] Xiaogang Xu [2] Gim Hee Lee [1]

[1] Department of Computer Science, National University of Singapore
[2] The Chinese University of Hong Kong

taohu@nus.edu.sg yihouxiang@gmail.com yan.zhiwen@nus.edu gimhee.lee@nus.edu.sg

Abstract

This paper presents Particle Rendering (PR), a new implicit rendering approach that extends Neural Radiance Fields (NeRF) by incorporating incident light along with traditional outgoing light modeling. In our framework, a 3D scene consists of a mass of particles, each offering a deeper understanding of light interactions by reflecting and emitting light in all directions. Our methodology involves a three-phase training pipeline: 1) Estimating the outgoing light field through a NeRF model; 2) Distilling the incident light field. A simple metric is introduced to assess the quality of the ray for better supervision; 3) Implicit rendering. We propose an implicit method to aggregate incident and outgoing fields that leverages Multilayer Perceptrons (MLP) to directly infer final pixel values, thus avoiding the limitation of traditional physically-based rendering techniques. The effectiveness of PR is demonstrated through state-of-the-art results in various challenging indoor and outdoor scenes, emphasizing its capability to handle complex lighting and reflective materials.

1. Introduction

Recently, remarkable advances have been made in the field of implicit functions with regard to computer vision and graphics, resulting in the generation of point-level scene comprehension and the production of realistic images in the form of novel view synthesis. Among these techniques, a pioneering method, called Neural Radiance Fields (NeRF) [17], has become increasingly popular due to its ability to synthesize ground-breaking high-quality images by optimizing implicit functions and volume rendering to map 3D geometry to color. Currently, NeRF-based methods have been widely applied to 3D visualization [3, 5, 8, 10, 29], 3D object reconstruction [9, 25, 28], and 3D AIGC [14, 21].

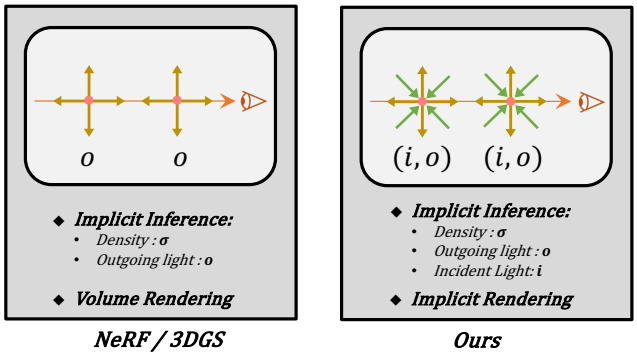


Figure 1. In comparison to NeRF's outgoing light (or radiance) field [17], our particle has an additional incident light field. Moreover, our PR adopt implicit rendering which enables us to render challenging scenes with more complex lighting and materials.

However, there are a number of reflective surfaces, such as glass, whose outgoing light is highly correlated to the incident light. Most existing implicit rendering approaches [12, 17] focus only on modeling the emitted light (or radiance) from each point of the scene but omits the effect of incident light. This limitation is especially noticeable when dealing with scenes with complicated lighting conditions and materials. For example, it is difficult to precisely represent and render an indoor scene with a large number of light sources and reflective mirrors. Some recent methods [6, 27, 31] combine Physically-based Rendering (PBR) [11], which uses the Bidirectional Reflectance Distribution Function (BRDF) to explain how materials interact with light. BRDF links outgoing light to incident light, providing a valuable connection to understand light transport. However, estimating BRDFs in a complex scene requires the impractical acquisition of precise normal directions, which can be difficult and computationally expensive for scene-level applications when only multi-view images can be accessed.

In this paper, we introduce Particle Rendering (PR), a novel implicit rendering approach that overcomes the limitations of existing methods when dealing with lighting and materials during rendering. Specifically, we represent a 3D scene with a mass of particles in our framework. Our representation is geometrically equivalent to the sampled points in NeRF, while each particle can receive incident light, apart from emitting outgoing light from and to all directions. Particle representation allows for a more comprehensive understanding of light transport in intricate environments. Our proposed PR builds upon the success of NeRF and advances by extending its implicit rendering paradigm. Instead of focusing solely on modeling the outgoing light distribution, our PR optimizes neural networks to map scene parameters to both outgoing and incident light information. This makes our PR a promising framework for generating realistic images in challenging scenarios. The essential differences between NeRF and ours are illustrated in Fig. 1.

Our PR follows a **three-phase training pipeline** designed to enhance the rendering quality of novel views. The **first phase** is dedicated to estimating the outgoing light field. A NeRF model is optimized to initially fit the density and outgoing light fields. This is accomplished using only multiple posed images as input data, relying on the power of implicit functions and volume rendering [15]. Consequently, we lay the foundation for a comprehensive understanding of the scene elements and their interaction with light. The **second phase** of our pipeline focuses on distilling the incident light field. Building on the knowledge acquired during the first phase, where the density and outgoing parameters of each particle were obtained, we now possess the essential information required to calculate the incident light from all possible directions for each particle. This is achieved through the utilization of the volumetric rendering capabilities of NeRF. We treat the rendered incident light as the teacher and the one using implicit function as the student, which allows us to distill the incident light environment for each particle. In the **final phase**, the incident and outgoing light fields are implicitly aggregated. This process enables us to render an image that faithfully captures the intricacies of the original environment by encompassing the collective interplay of light rays as they traverse the scene and hit the particles. Through this three-phase training pipeline, our PR emerges as a powerful framework for generating photo-realistic images in the real world.

There are two essential problems that need to be addressed during the three phases of our framework. **1) Quality of Distillation**. In the first phase, only the outgoing directions that point towards the camera center are fully supervised, and thus leading to the under-fitting problem for outgoing rays along other directions. This affects the quality of incident light and subsequent distillation. This is also a limitation of previous NeRF-based methods. To circumvent the under-fitting issue, we propose a metric in the second phase to assess the quality of each incident ray and assign the appropriate weights to distill the incident light field. Our motivation is that high-quality incident rays should have a concentrated distribution considering their rendered weights. For example, the quality of the ray is usually poor when the variance in the weight distribution of an incident ray is large [1, 2]. **2) Strategy of Aggregation**. During the final stage, adopting the conventional PBR method to aggregate the incident and outgoing light field involves predicting the directions of surface normal and material parameters and then employing conventional graphics-based rendering to determine the color of outgoing light. This complicates the process and significantly increases the computational complexity as we only aim to predict the final color without reconstructing additional physical parameters. Instead of relying on traditional PBR methods to connect outgoing and incident light, our approach takes an implicit route. It just combines the embeddings of the incident and outgoing light from the particles, and uses an MLP to directly calculate the final RGB values for each ray. This approach improves computational efficiency and accessibility, eliminating the need to estimate physical parameters, and thus achieves better performance.

Contributions. Our contributions can be summarized as:

1. We propose a Particle Rendering to take account of incident light during novel-view image rendering. It extended the outgoing paradigm of NeRF in favor of a scene of complex lighting and reflective materials.
2. We introduce a metric to evaluate the quality of incident rays, improving the distillation of incident light fields.
3. We propose an implicit while effect method to aggregate the incident light field and outgoing light field to render the final pixel colors.
4. Experiments conducted on various datasets demonstrate that our PR is capable of achieving state-of-the-art performance in both indoor and outdoor datasets and scenes with challenging lighting and materials.

2. Related Work

Our work is related to previous research on implicit function in 3D vision, NeRF-based Novel View Synthesis, and rendering complex lighting effects.

Implicit Function in 3D Vision. The implicit function in 3D computer vision represents a neural mapping from 3D geometry that includes location and direction to associated attributes such as shape and appearance. The concept is initially explored in the Occupancy Network [16], which utilizes an MLP to determine the occupancy status of a point based on its location. However, this method requires ground-truth point-occupancy pairs that are costly to annotate. Subsequently, DVR [20] introduces training of

implicit functions through differentiable rendering modules, but is mainly confined to object-level performance. Further developments in implicit rendering involved the use of Convolutional Networks (ConvNet) [13] or Transformer [23] to synthesize images directly [19, 33], bypassing complex physical rendering processes. However, these methods often overfit to training images or views, limiting their generalization. In our work, we follow the principle of implicit function and propose the integration of both incident and outgoing light into a rendering pipeline. Our approach is particularly effective for realistically rendering real-world scenes with reflective properties, marking a significant step forward in the field of implicit rendering.

NeRF-based Novel View Synthesis. NeRF [17] utilizes MLPs to compute color and density in continuous 5D coordinates, enhanced through differentiable volume rendering [15] based on input views. Researchers have made several advances to refine NeRF, focusing on image quality and rendering efficiency. Mip-NeRF [1] improves fine-detail rendering and anti-aliasing through prefiltering of input images. Mip-NeRF 360 [2] extends the concept to 360-degree scene rendering, effectively utilizing input images covering the entire environment. Zip-NeRF [3] combines Mip-NeRF 360 [2] with grid-based neural radiance fields (as in Instant-NGP), with the aim of improving performance and accelerating training. Nerfacto [22] integrates the strengths of various NeRF methodologies such as hash encoding [18], fast MLP inference, and advanced regularization of Mip-NeRF 360 [2]. Despite the advances, these state-of-the-art methods have not fully addressed the impact of environmental lighting and surface materials simultaneously, making it challenging to synthesize scenes with reflective qualities accurately. This gap indicates a potential area for future research and development in NeRF technology, with the aim of producing even more realistic and photorealistic renderings in challenging lighting and material scenarios.

Rendering Complex Lighting. Achieving photorealistic rendering typically involves replicating complex lighting interactions and material properties. Classical approaches utilize rasterization or ray tracing with analytical BRDFs for this purpose. NeILF [27] represents a significant advance for rendering high-quality images, particularly for reflective objects, but requires pre-existing ground-truth mesh data. To address this limitation, NeILF++ [31] is developed for the object-level, which can predict both the surface and the appearance at the same time. Ref-NeRF [24] incorporates diffuse and specular color relative to surface normal in the rendering process, substantially enhancing the rendering quality of scenes with reflective materials. However, both NeILF++ and Ref-NeRF struggle with scene-level rendering during application because due to the inaccuracies in getting so many surface normals and various materials. Mirror-NeRF [30] from Zeng *et al.* took on the challenge of rendering scenes containing mirrors, a task complicated by the difficulty in discerning the inside and outside of mirrors to create photorealistic images. Unlike previous methods, Mirror-NeRF required access to the ground-truth mask of the mirrors. In contrast, our approach simplifies the process by requiring only input images and avoiding predicting additional physical parameters.

In summary, compared with previous work on novel view synthesis using implicit functions, our particle rendering framework extends the NeRF-based paradigm to address detailed lighting and material effects in rendered images. We distinguish ourselves as the first to concurrently model incident and outgoing radiance within an implicit neural framework, marking a significant advancement in the field of photorealistic rendering.

3. Approach

In this paper, our goal is to design a renderer that is sensitive to incident and outgoing light and can generate high-fidelity images from novel views, given multi-view images with calibrated cameras. In this section, we elaborate on the details of our proposed framework. As outlined previously, our PR consists of three phases: estimating outgoing light field, distilling incident light field, and implicit rendering.

Definition of a Particle. Formally, we define a particle as one of the minimal units that receives and emits light in a 3D scene. For simplicity, it is geometrically equal to a point sampled in the NeRF [17], which usually indicates the points around the surface of the objects. The same multi-stage sampling strategy as NeRF is adopted to select particles for each ray. In contrast to the sampled points in NeRF that only emit outgoing light, our particle takes into account both incident and outgoing light during rendering.

3.1. Outgoing Light Field

The first phase focuses on optimizing implicit functions to predict the outgoing light color $\mathbf{o}$ of any particle at 3D location $\mathbf{x}$ and direction $\mathbf{d}$, which is also the objective of NeRF [17]. Since vanilla NeRF is not efficient in both training and rendering compared to subsequent methods [5, 8, 10, 22, 29], we adopt the recently widely used method called Nerfacto [22] as our backbone. Nerfacto shares the same principle and pipeline as NeRF, but it is significantly more efficient in representing the scene as continuous implicit functions with the introduction of more advanced techniques [2, 18]. These implicit functions map a 3D coordinate $\mathbf{x}$ and a 2D direction $\mathbf{d}$ to the density σ and the outgoing light $\mathbf{o}$, *i.e.*:

$$\sigma, \mathbf{e_o} = \mathcal{F}_o(\mathbf{x}), \quad \mathbf{o} = \mathcal{O}(\mathbf{e_o}, \mathbf{d}) \tag{1}$$

where $\mathbf{e_o}$ is the embedding for outgoing light that represents the extracted feature. $\mathcal{F}_o(\mathbf{x})$ and $\mathcal{O}(\mathbf{e_o}, \mathbf{d})$ are trained by

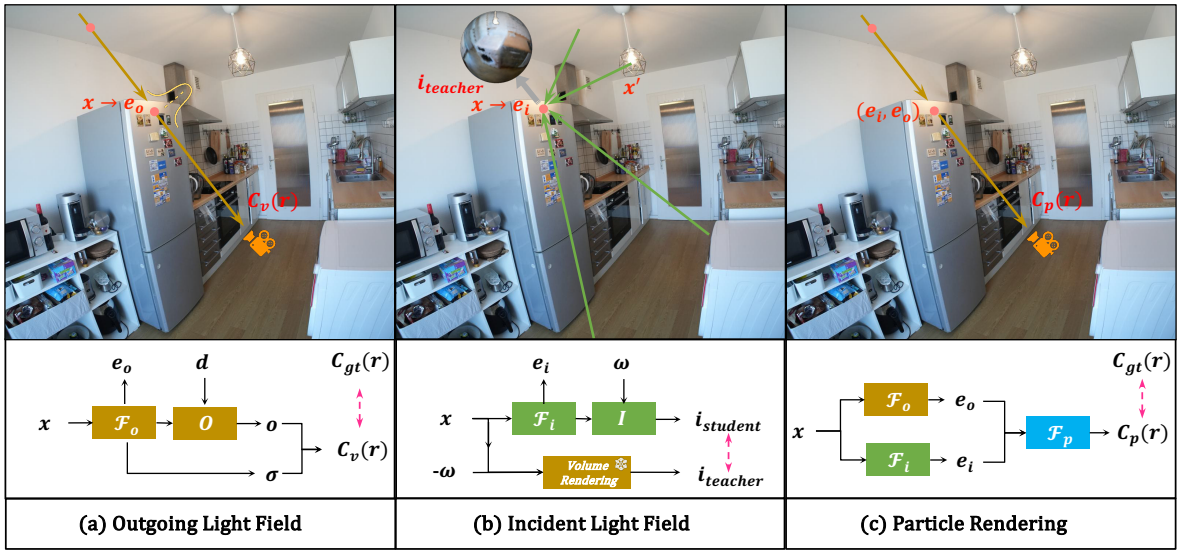


Figure 2. Overview of our proposed Particle Rendering. There are three phases: (a) **Outgoing Light Field.** we optimize a neural outgoing light field using the principle of NeRF [17], which allows us to predict the initial emitted light color and embedding of all particles. (b) **Incident Light Field.** We then distill the incident light and embedding out via the frozen volume rendering of (a). (c) **Implicit Rendering.** We concatenate the basic light embeddings e_i and e_o as representations of the ray and render them to the final color of the pixels. The brown and green arrows indicate the outgoing and incident direction, respectively.

minimizing the volume rendering loss between the rendered and input pixel colors. Specifically, for each camera ray $\mathbf{r}$, we sample N particles along the ray and accumulate the color $\mathbf{C}_v(\mathbf{r})$ using volume rendering:

$$\begin{aligned} \mathbf{C}_v(\mathbf{r}) &= \sum_{n=1}^{N} w_n \mathbf{o}_n, \\ w_n &= T_n(1 - \exp(-\sigma_n \delta_n)), \\ T_n &= \exp(-\sum_{j=1}^{n-1} \sigma_j \delta_j), \end{aligned} \tag{2}$$

where T_n is the transmittance, δ_n is the distance between the sample particles, w_n indicates the weight during color rendering, and $\sigma_n, \mathbf{o}_n$ are the predicted density and outgoing light color in the sample location $\mathbf{x}$. The loss function is defined as:

$$\mathcal{L}_o = \sum_{\mathbf{r} \in \mathcal{R}} ||\mathbf{C}_v(\mathbf{r}) - \mathbf{C}_{gt}(\mathbf{r})||^2, \tag{3}$$

where $\mathcal{R}$ denotes the set of rays casted from camera and $\mathbf{C}_{gt}(\mathbf{r})$ is the ground truth color. Through optimizing this loss, the implicit functions learn to map spatial locations and viewing directions to outgoing light for novel views.

3.2. Incident Light Field

The next phase aims to predict the incident light $\mathbf{i}$ of each particle at location $\mathbf{x}$ along the incident direction ω. Inspired by the outgoing radiance field, we adopt implicit functions $\mathcal{F}_i$ and $\mathcal{I}$ to obtain the incident embedding $\mathbf{e_i}$ and incident light $\mathbf{i}$, respectively:

$$\mathbf{e_i} = \mathcal{F}_i(\mathbf{x}), \quad \mathbf{i} = \mathcal{I}(\mathbf{e_i}, \omega). \tag{4}$$

Distillation. To guide the learning process, the rendered color $\mathbf{C}_v(t)$ along incident ray is treated as the teacher $\mathbf{i}_{teacher}$, where $\mathbf{t} = (\mathbf{x}, -\omega)$ denotes the ray at the same particle location while in the opposite direction of incident light $-\omega$. The predicted incident light $\mathbf{i}$ plays the role of the student $\mathbf{i}_{student}$. In other words, we can distill the incident light field from the learned outgoing light field. This reference is established via volume rendering as Eq. 2. We introduce the distillation loss $\mathcal{L}_i$, defined by:

$$\begin{aligned} \mathcal{L}_i &= \mathcal{M}(\mathbf{t}) \cdot ||\mathbf{i}_{teacher} - \mathbf{i}_{student}||^2 \\ &= \mathcal{M}(\mathbf{t}) \cdot ||\mathbf{C}_v(\mathbf{t}) - \mathbf{i}||^2. \end{aligned} \tag{5}$$

Here, $\mathcal{M}(\mathbf{t})$ signifies the importance assigned to the incident ray $\mathbf{t}$, and the loss constitutes the squared square error between the predicted and referenced rendering color.

Incident Direction. The indiscriminate sampling of particles across all spatial and angular coordinates frequently yields sub-optimal results, a consequence of the generalization limitations of volume rendering for significantly different locations and directions. We thus anticipate that the incident direction of the target particle $\mathbf{x}$ should point to

another randomly selected particle $\mathbf{x}'$ rather than towards a random point in the scene, *i.e.*:

$$\omega = \frac{\mathbf{x} - \mathbf{x}'}{||\mathbf{x} - \mathbf{x}'||_2}. \tag{6}$$

Selecting these incident directions is more efficient because the incident light of a particle is exactly from the outgoing light of another group of particles.

Incident Weight. Since the outgoing light field is optimized only on observed camera rays, its rendering quality limits across various other positions and orientations. To address this under-fitting issue, we introduce a metric, called *Certainty Score*, for each incident light, assessing the reliability of rendered rays, and refining the distillation process. The metric for an incident ray $\mathbf{t}$ quantifies the confidence coefficient in rendering the teacher $\mathbf{i}_{teacher}$, formulated as:

$$\mathcal{M}(\mathbf{t}) = \sum_{n=1}^{N} w_n^2. \tag{7}$$

The weight of the n-th particle during volume rendering, w_n, is expressed in Eq. 2. It is easy to see that $\mathcal{M}(\mathbf{t})$ is in the range of $[0, 1]$. This proposed certainty score is then used as the weight of the incident rays in Eq. 5. This metric is straightforward yet effective in improving the quality of distillation. There is a close linear relationship between $\mathcal{M}(\mathbf{t})$ and the variance of w_n. Therefore, a higher score implies a more concentrated weight distribution, which suggests a more reliable rendering color.

3.3. Implicit Rendering

The final phase aggregates the estimated incident and outgoing light fields to produce the rendered images. Previous approaches such as NeILF [27] usually use a physically-based rendering to transform the outgoing light to incident light through the Disney BRDF [4]. In computer graphics, the explicit rendering equation [11] that considers incident light is represented as:

$$\mathbf{o}' = \sum_{\omega \in \mathbf{\Omega}} b(\mathbf{x}, \mathbf{r}, \omega) \mathcal{I}(\mathcal{F}_i(\mathbf{x}), \omega)(\omega \cdot \mathbf{n}), \tag{8}$$

where $\mathbf{\Omega}$ is the set of incident directions for particle $\mathbf{x}$, $\mathbf{o}'$ is the physical outgoing light, $b(\mathbf{x}, \mathbf{r}, \omega)$ is the BRDF value for incident direction ω and outgoing direction $\mathbf{r}$, and $\mathbf{n}$ is the normal direction.

The difficulty is that we must first predict precise normal directions and BRDF material parameters, which is challenging at the scene level [31]. Furthermore, computing the incident light from all possible directions takes too much time and memory. Since we have already captured all shape and appearance information of the particle via the incident and outgoing embedding, it becomes apparent to fuse them and implicitly predict the final image colors via an MLP.

Consequently, we propose the final implicit function as our particle renderer, which directly combines incident and outgoing light embeddings to predict RGB colors, skipping explicit reconstruction of physical properties. Specifically, for each particle sampled along a camera ray, we extract embeddings $\mathbf{e}_\mathrm{i}, \mathbf{e}_\mathrm{o}$ from incident and outgoing light fields using the encoder networks $\mathcal{F}_i, \mathcal{F}_o$. The embeddings are then concatenated and fed into an MLP $\mathcal{F}_p$ to predict ray color:

$$\mathbf{C}_p(\mathbf{r}) = \mathcal{F}_p(\mathbf{e}_i^1, \mathbf{e}_o^1, \mathbf{e}_i^2, \mathbf{e}_o^2, ..., \mathbf{e}_i^N, \mathbf{e}_o^N). \tag{9}$$

The loss function of implicit rendering is the mean square error between the rendered color and the ground truth given by:

$$\mathcal{L}_p = \sum_{\mathbf{r} \in \mathcal{R}} ||\mathbf{C}_p(\mathbf{r}) - \mathbf{C}_{gt}(\mathbf{r})||^2. \tag{10}$$

Our implicit function for rendering provides a learnable mapping from incident and outgoing light to final pixel colors tailored to the scene. Intuitively, the model learns to select and combine relevant illumination and material interactions implicitly for accurate view synthesis. Our aggregation approach has several benefits. First, it avoids challenging and time-consuming intermediate predictions of normals and BRDFs. Second, the embeddings provide bottlenecks that compress the 4D radiance fields into informative global-scene representations. Finally, the implicit model learns specialized mappings per scene that are adapted to lighting and materials.

Training Losses. The final training goal is optimizing the following equation:

$$\mathcal{L} = \lambda_o \cdot \mathcal{L}_o + \lambda_i \cdot \mathcal{L}_i + \lambda_p \cdot \mathcal{L}_p, \tag{11}$$

where λ_o, λ_i, and λ_p controls the weights of outgoing, incident and implicit rendering loss respectively. The implicit functions that need to be trained are $\mathcal{F}_o, \mathcal{O}, \mathcal{F}_i, \mathcal{I}$, and $\mathcal{F}_p$.

Our PR pipeline provides a novel solution to model complex light transport for high-fidelity view synthesis. The outgoing radiance field captures geometric and emissive properties. The incident field represents the incident illumination. Finally, the aggregator combines these to render images with realistic lighting effects suitable for scenes with complex material and lighting environment.

4. Experiments

4.1. Experimental Setup

Datasets. We conduct experiments to evaluate our PR framework against existing state-of-the-art approaches on the indoor and outdoor datasets: *360 Dataset* [1], and our collected *360 reflective Dataset* which contains various intricate lighting environments and reflective materials.

Evaluation Metrics. We evaluate the quality of the synthesized images by reporting Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) [26], and Learned Perceptual Image Patch Similarity (LPIPS) [32] metrics. These metrics compare the rendered images with the ground truth, where higher PSNR and SSIM, as well as lower LPIPS, signify superior synthesis quality.

Implementation Details. All implicit functions are implemented by MLPs. During the estimation of the outgoing light field, the number of particles per ray N is set to 32. For incident light field distillation, all rendered colors along incident directions via volume rendering are detached, and no gradient is propagated back to outgoing implicit functions. The maximum number of training iterations is 70,000 with a batch size of 4096 for all three phases. At 5,000 iterations, the distillation and implicit rendering phases are inserted simultaneously for joint training. We set the loss weight λ_o, λ_i, and λ_p to $1, 0.1$, and 1, respectively. The training process for each scene is completed within one hour and requires only 8GB of GPU memory. Refer to our supplementary for additional network and dataset evaluation details with regard to the experiments.

4.2. Comparisons

Comparison on Mip-NeRF 360 Dataset. The *Mip-NeRF 360 Dataset* [1] is popular for its high-resolution images and exceptional quality, capturing comprehensive 360-degree views of the real world. This dataset has become a benchmark in the research field, with many recent methods, such as Mip-NeRF [1], Mip-NeRF 360 [2], and Zip-NeRF [3] utilizing it to evaluate performance. There are a total of 7 public-released scenes: Bicycle, Garden, Stump, Room, Counter, Kitchen and Bonsai. We adopt the standard train and test split. The experimental results of ours and the baselines are presented in the Tab. 1. Ours achieves the best performance, with a 0.47dB improvement compared to the state-of-the-art Zip-NeRF [3]. This superiority underscores the importance of modeling incident light, particularly in complex environments with challenging materials. Moreover, compared to the Outgoing rendering results (Stage 1), our method achieves significantly better performance. Additionally, we implemented the Outgoing network with the same number of parameters (9 M) as the Implicit rendering (noted as Outgoing, Large Model), the performance only slightly improved over the original Outgoing results. **This indicates that the enhanced performance of our method is not due to an increase in parameters.** The comparative under-performance of other methods highlights a crucial insight: mere extensions of NeRF are not adequate to capture the nuanced lighting of real-world settings. Note that Nerfacto [22] runs without pose optimizer and appearance embedding for best evaluation performance.

Comparison on 360 Reflective Dataset. To further underscore the capabilities of our approach, we conduct extensive experiments in our collected dataset: *360 reflective Dataset*. This dataset which poses greater challenges due to its complexity, comprises eight real-world scenes characterized by intricate lighting and diverse materials. These scenes feature elements such as multiple light sources, reflective flooring, and mirrors. Among them, four scenes are sourced from the ScanNet dataset [7], a large-scale resource tailored for indoor settings that integrates high-quality color imagery with economical geometry. One scene is selected from the Nerfstudio [22] dataset. Additionally, two derived from Mirror-NeRF [30] are notable for the prominent inclusion of large mirrors in the room centers, while the remaining one scene is captured by our own. There lacks a fixed image resolution and number for each scene of the dataset, with resolutions ranging from 640×480 to $1,920 \times 1,080$ and numbers from 100 to 600. We employ the same split as the *Mip-NeRF 360 Dataset* to every 8th frame for testing and the rest for training. The results of the experiment are detailed in Tab. 1. Ours achieved significantly higher performance than all state-of-the-art methods (1.56 dB improvement), which demonstrates the substantial benefits of considering both incident and outgoing light fields in complex environments.

Qualitative Comparisons. We also provide a qualitative demonstration of the performance of our proposed PR method compared to other baseline approaches in both indoor and outdoor scenes. These comparative results are illustrated in Fig. 3 and Fig. 4. A key strength of our PR lies in its exceptional sensitivity to intricate lighting conditions and a diverse range of materials. This sensitivity is particularly evident in environments where complex interplays of light are present or where materials with unique reflective or absorptive properties are encountered. For example, in the Poster scene of Fig. 4, there is a glass which reflects the background light when viewing in different directions. Nerfacto [22] cannot properly deal with the case, thus rendering a plain light effect. Ref-NeRF [24] also does not perform well due to their strict requirement of surface normal. . Compared to other existing methods, PR consistently exhibits superior performance and is adept at handling scenarios that pose significant challenges due to their complexity of lighting and material diversit. Ours successfully renders the enriching light in the surface of the reflective objects. This performance advantage is a testament to the efficacy of our approach in accurately simulating the intricacies of real-world lighting and material interactions, setting a new standard in the field of particle rendering.

4.3. Ablation Studies

There are a variety of different approaches to achieve our incident distillation and light aggregation. In the following experiments, we analyze the influence of different modules

Table 1. Comparison on the *Mip-NeRF 360* dataset and our collected *Reflective 360* dataset. All evaluated scenes contain complicated reflective surfaces or challenging lighting environments.

Methods	*MipNeRF 360*			*Reflective 360*		
	PSNR (↑)	SSIM (↑)	LPIPS (↓)	PSNR (↑)	SSIM (↑)	LPIPS (↓)
Instant-NGP [18]	23.94	0.641	0.478	23.02	0.824	0.398
NeRF [17]	24.85	0.659	0.426	23.72	0.837	0.342
Mip-NeRF [1]	25.12	0.672	0.414	26.98	0.834	0.274
Nerfacto [22]	27.98	0.800	0.291	26.91	0.846	0.288
3D Gaussian Splatting [12]	28.96	0.859	0.186	27.62	0.857	0.260
Ref-NeRF [24]	28.35	0.843	0.218	27.86	0.864	0.247
Mip-NeRF 360 [2]	29.23	0.846	0.208	27.44	0.855	0.262
Zip-NeRF [3]	30.08	0.877	0.170	28.63	0.868	0.235
Ours (Outgoing)	28.00	0.809	0.279	27.26	0.851	0.288
Ours (Outgoing, Large Model)	28.14	0.821	0.270	27.33	0.854	0.275
Ours (Outgoing + Incident)	30.55	0.891	0.150	30.19	0.890	0.202

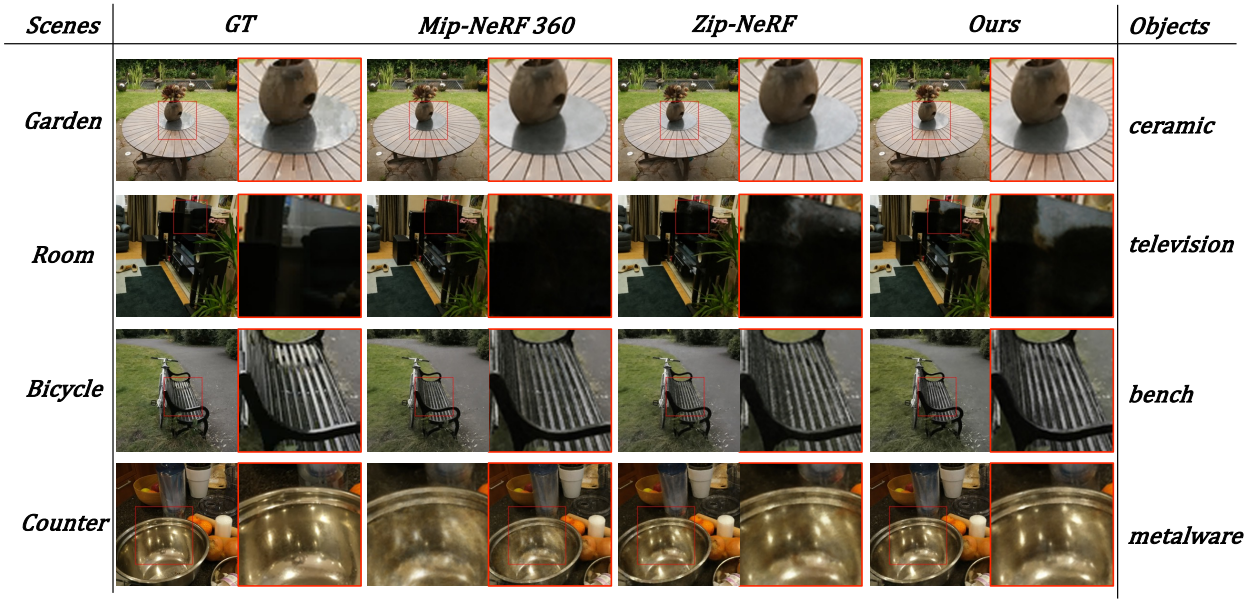


Figure 3. Qualitative comparisons on the *360 Dataset*. We visualize two indoor scenes and two outdoor scenes, including the Garden, Room, Bicycle, and Counter. We selected and enlarged reflective objects (ceramic, television, bench, and metalware) for better demonstration.

to prove the efficacy of our proposed solutions for obtaining incident light fields and combining them into the final images. The experimental results are presented in Tab. 2.

The Effect of Incident Directions. Determining the incident directions of the particles presents multiple options. A naive approach is to randomly select directions in all 360 degree directions despite low effectiveness, as shown in Fig. 5(a). Physically-based methods typically first obtain the normal direction of the surface [31] and then sample the directions within the corresponding hemisphere. However, this approach faces two significant challenges: it is difficult to accurately predict normal directions, as illustrated in Fig. 5(b). In contrast, our proposed method utilizes particle-to-particle incident directions as Eq. 6. This strategy eliminates the need to predict additional physical parameters and improve the effective of distillation, as shown in Fig. 5(c). We compared these two strategies to determine the incident direction on the *Reflective 360 dataset* as shown in Row 1 $\sim$ 3, Tab. 2. Random incidents yield the poorest results, while physical incidents are slightly better. In comparison, our particle-to-particle incidents yield the best outcomes.

The Effect of Incident Weights. We conducted an experiment in which we eliminated the incident weighting to assess the impact of incident weights. The experiment results in Row 3 of Tab. 2 shows that assigning a weight to each incident ray using Eq. 7 can help to filter out uncertain cases

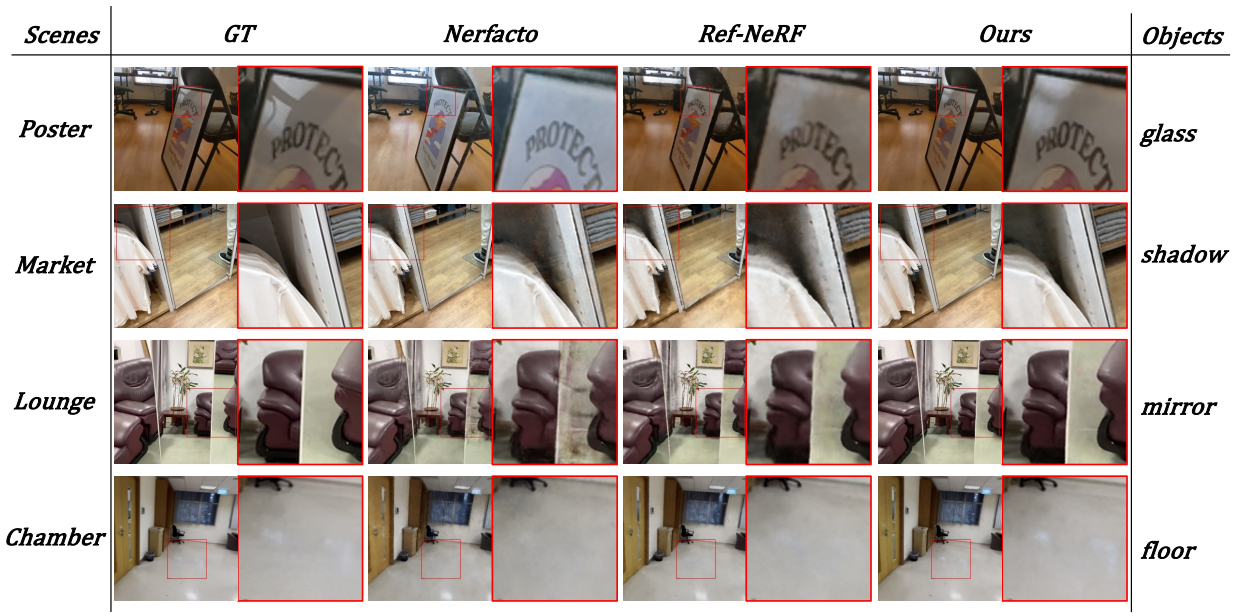
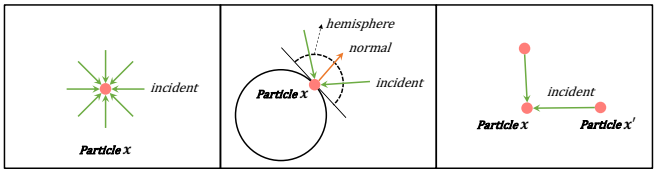


Figure 4. Qualitative comparisons on the *Reflective 360 Dataset*. We visualize four indoor scenes, including the Poster, Market, Lounge, and Chamber. We selected and enlarged the reflective objects (glass, shadow, mirror, and floor) for a better demonstration.



(a) Random Incident (b) Physical Incident (c) Particle-to-Particle Incident (Ours)

Figure 5. Incident directions of random, physically-based and our particle-to-particle strategies. The normal direction of the former must be predicted initially, whereas our approach does not require this step and is more efficient.

during distillation, resulting in an improvement of the PSNR metric by 1.28 dB.

The Effect of Implicit Rendering. Finally, we validate the importance of implicit rendering, incident light field, and outgoing light field. We remove all outgoing embeddings $\mathcal{F}(\mathbf{e}_i^1, \mathbf{e}_i^2, ..., \mathbf{e}_i^N)$ or incident embeddings $\mathcal{F}(\mathbf{e}_o^1, \mathbf{e}_o^2, ..., \mathbf{e}_o^N)$ to render novel views. An interesting observation is that implicit rendering with only outgoing incident embeddings also achieved satisfactory performance in Row 6 of Tab. 2, which demonstrates the potential of implicit function in rendering novel-view images. We implement an explicit rendering, following NeILF++ [31], which additionally predicts the normal directions and BRDF parameters. However, the PSNR is only 27.25 dB, demonstrating the difficulty in predicting physical parameters. For explicit rendering, we sample 128 incident light for each particle, GPU memory usage reaches 20 GB, and the rendering speed for image resolution at 960×540 is around 10 seconds per image, which is $20\times$ slower compared to our implicit rendering at 2 FPS.

No.	Method	PSNR ↑	SSIM ↑	LPIPS ↓
1	PR	**30.19**	**0.890**	**0.202**
2	PR w/ random incidents	28.02	0.863	0.265
3	PR w/ physical incidents	28.57	0.870	0.247
4	PR w/o incident weighting	28.91	0.873	0.246
5	PR w/o outgoing field	28.01	0.854	0.274
6	PR w/o incident field	28.83	0.865	0.236
7	PR w/ explicit rendering [31]	27.25	0.852	0.283

Table 2. Ablation studies on *Reflective 360 Dataset* to validate the effect of incident direction, weighting and implicit rendering.

5. Conclusion

In this paper, we introduce a new implicit rendering framework for novel view synthesis, which is especially suitable for complex lighting environments and multiple reflective materials. Instead of attempting to create a more accurate representation of the physically-based rendering process, we follow the trend of implicit functions in 3D vision and proposed implicit functions for rendering images. We first extract the outgoing light field using the NeRF, then distill the incident light field, and finally directly render the pixel color via the final implicit function. The results are surprisingly better than those of the state-of-the-art methods, demonstrating the importance of incident light during rendering and the potential of implicit rendering.

Acknowledgement. This research / project is supported by the National Research Foundation (NRF) Singapore, under its NRF-Investigatorship Programme (Award ID. NRF-NRFI09-0008).

References

[1] Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In: ICCV (2021) 2, 3, 5, 6, 7

[2] Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In: CVPR (2022) 2, 3, 6, 7

[3] Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Zip-nerf: Anti-aliased grid-based neural radiance fields. In: ICCV (2023) 1, 3, 6, 7

[4] Burley, B., Studios, W.D.A.: Physically-based shading at disney. In: Acm Siggraph (2012) 5

[5] Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H.: Tensorf: Tensorial radiance fields. In: ECCV (2022) 1, 3

[6] Chen, H., He, B., Wang, H., Ren, Y., Lim, S.N., Shrivastava, A.: NeRV: Neural representations for videos. In: NeurIPS (2021) 1

[7] Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Nießner, M.: ScanNet: Richly-annotated 3D reconstructions of indoor scenes. In: CVPR (2017) 6

[8] Fridovich-Keil and Yu, Tancik, M., Chen, Q., Recht, B., Kanazawa, A.: Plenoxels: Radiance fields without neural networks. In: CVPR (2022) 1, 3

[9] Ge, W., Hu, T., Zhao, H., Liu, S., Chen, Y.C.: Ref-neus: Ambiguity-reduced neural implicit surface learning for multi-view reconstruction with reflection. In: ICCV (2023) 1

[10] Hu, T., Liu, S., Chen, Y., Shen, T., Jia, J.: Efficientnerf efficient neural radiance fields. In: CVPR (2022) 1, 3

[11] Kajiya, J.T.: The rendering equation. Proceedings of the 13th annual conference on Computer graphics and interactive techniques (1986), https://api.semanticscholar.org/CorpusID:9226468 1, 5

[12] Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4) (July 2023), https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/ 1, 7

[13] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE (1998) 3

[14] Liu, R., Wu, R., Hoorick, B.V., Tokmakov, P., Zakharov, S., Vondrick, C.: Zero-1-to-3: Zero-shot one image to 3d object (2023) 1

[15] Max, N.: Optical models for direct volume rendering. IEEE TVCG (1995) 2, 3

[16] Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy networks: Learning 3d reconstruction in function space. In: CVPR (2019) 2

[17] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: ECCV (2020) 1, 3, 4, 7

[18] Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM TOG (2022) 3, 7

[19] Nguyen-Phuoc, T., Li, C., Theis, L., Richardt, C., Yang, Y.L.: Hologan: Unsupervised learning of 3d representations from natural images. In: ICCV (2019) 3

[20] Niemeyer, M., Mescheder, L.M., Oechsle, M., Geiger, A.: Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In: CVPR (2020) 2

[21] Poole, B., Jain, A., Barron, J.T., Mildenhall, B.: Dreamfusion: Text-to-3d using 2d diffusion. arXiv (2022) 1

[22] Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Kerr, J., Wang, T., Kristoffersen, A., Austin, J., Salahi, K., Ahuja, A., McAllister, D., Kanazawa, A.: Nerfstudio: A modular framework for neural radiance field development. In: ACM SIGGRAPH 2023 Conference Proceedings (2023) 3, 6, 7

[23] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I.: Attention is all you need. In: NeurIPS (2017) 3

[24] Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T., Srinivasan, P.P.: Ref-NeRF: Structured view-dependent appearance for neural radiance fields. In: CVPR (2022) 3, 6, 7

[25] Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In: NeurIPS (2021) 1

[26] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE TIP (2004) 6

[27] Yao, Y., Zhang, J., Liu, J., Qu, Y., Fang, T., McKinnon, D., Tsin, Y., Quan, L.: Neilf: Neural incident light field for physically-based material estimation. In: ECCV (2022) 1, 3, 5

[28] Yariv, L., Gu, J., Kasten, Y., Lipman, Y.: Volume rendering of neural implicit surfaces. In: NeurIPS (2021) 1

[29] Yu, A., Li, R., Tancik, M., Li, H., Ng, R., Kanazawa, A.: PlenOctrees for real-time rendering of neural radiance fields. In: ICCV (2021) 1, 3

[30] Zeng, J., Bao, C., Chen, R., Dong, Z., Zhang, G., Bao, H., Cui, Z.: Mirror-nerf: Learning neural radiance

fields for mirrors with whitted-style ray tracing. In: ACMMM (2023) 3, 6

[31] Zhang, J., Yao, Y., Li, S., Liu, J., Fang, T., McKinnon, D., Tsin, Y., Quan, L.: Neilf++: Inter-reflectable light fields for geometry and material estimation. In: ICCV (2023) 1, 3, 5, 7, 8

[32] Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR (2018) 6

[33] Zhang, Y., Chen, W., Ling, H., Gao, J., Zhang, Y., Torralba, A., Fidler, S.: Image gans meet differentiable rendering for inverse graphics and interpretable 3d neural rendering. In: ICLR (2021) 3