SpectralNeRF: Physically Based Spectral Rendering with Neural Radiance Field

Ru Li[1], Jia Liu[2], Guanghui Liu[2*], Shengping Zhang[1] , Bing Zeng[2], Shuaicheng Liu[2*]

[1]Harbin Institute of Technology, Weihai, China
[2]University of Electronic Science and Technology of China, Chengdu, China
{liru, s.zhang}@hit.edu.cn, {liujia21@std., guanghuiliu@, eezeng@, liushuaicheng@}uestc.edu.cn

Abstract

In this paper, we propose SpectralNeRF, an end-to-end Neural Radiance Field (NeRF)-based architecture for high-quality physically based rendering from a novel spectral perspective. We modify the classical spectral rendering into two main steps, 1) the generation of a series of spectrum maps spanning different wavelengths, 2) the combination of these spectrum maps for the RGB output. Our SpectralNeRF follows these two steps through the proposed multi-layer perceptron (MLP)-based architecture (SpectralMLP) and Spectrum Attention UNet (SAUNet). Given the ray origin and the ray direction, the SpectralMLP constructs the spectral radiance field to obtain spectrum maps of novel views, which are then sent to the SAUNet to produce RGB images of white-light illumination. Applying NeRF to build up the spectral rendering is a more physically-based way from the perspective of ray-tracing. Further, the spectral radiance fields decompose difficult scenes and improve the performance of NeRF-based methods. Comprehensive experimental results demonstrate the proposed SpectralNeRF is superior to recent NeRF-based methods when synthesizing new views on synthetic and real datasets. The codes and datasets are available at https://github.com/liru0126/SpectralNeRF.

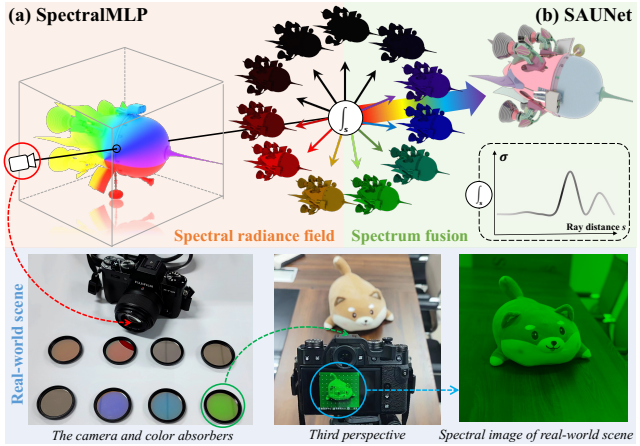


Figure 1: SpectralNeRF builds up the process of spectral rendering. The first step samples spectrum points along the ray and uses volume rendering to generate spectrum maps. The second step fuses discrete spectrum maps to obtain RGB images. *Up:* our pipeline. *Bottom:* our capture device.

Introduction

Newton found that white light can be dispersed into a series of spectrums with various colors from red to violet by passing light through the glass prism (Newton 1672). After that, the research on spectral theory developed rapidly (Pickholtz, Schilling, and Milstein 1982; Helffer 2013). Until now, the application of spectral images has involved various aspects of daily life, including object detection (Liang et al. 2018), face recognition (Uzair, Mahmood, and Mian 2015), and so on. Spectral images can record and reveal the electromagnetic radiation intensity information of objects, which is an important interdisciplinary subject mainly involving physics and chemistry (Bertrand et al. 2021).

Spectral rendering is a fundamental problem in computer graphics, which can understand the absorption, reflection, and other interactions with objects and has been used to generate photo-realistic images (Peercy 1993). Conventional spectral rendering involves two transformations: 1) the spectral power distribution L to the XYZ image, achieved by using integral operation through the visible light; 2) the XYZ image to the RGB image, realized by the conversion matrix (Smits 1999). Based on the transformations of $L \rightarrow$ XYZ $\rightarrow$ RGB, physically-based spectral rendering has been widely researched over the past decades (Watanabe et al. 2013; Sun et al. 2001; Knaus and Zwicker 2011). Such methods predict photo-realistic rendering that no effect that contributes to the interaction of light with a scene is neglected. However, they merely generate one image of the current viewpoint and are limited in representing the scene.

Recently, Neural Radiance Field (NeRF) is designed to render compelling images of 3D scenes from novel viewpoints (Mildenhall et al. 2020). NeRF-based methods achieve photo-realistic rendering of scenes by encoding the volumetric density and color of a scene within the weights of a coordinate-based multi-layer perceptron (MLP). Subsequently, a series of works focused on recovering the radiance field using deep neural networks (Barron et al. 2021, 2022; Yan, Li, and Lee 2023). This approach has enabled signifi-

*Corresponding authors

cant progress toward photo-realistic view synthesis and can solve the limitation that spectral rendering cannot represent the overall scene. However, NeRF-based methods may lose important details when encountering complicated scenes.

In this paper, we propose a NeRF-based architecture to achieve the physically-based spectral rendering, named SpectralNeRF. We construct the spectral radiance field for the scene information along the ray, which facilitates the rendering process and keeps the rendering pipeline simple yet effective by employing a physically-based multi-spectral integral calculation for the $L \rightarrow \text{XYZ} \rightarrow \text{RGB}$ conversion. For spectral rendering, applying the neural radiance field to learn the spectrum maps and using the integration of spectral bands are more physically-based ways. For NeRF-based methods, complicated scenes can be simplified through multiple spectral components compared to only normal RGB images, which is superior to previous NeRF-based methods in terms of geometry and texture reconstruction under complex scenes and image quality of the novel viewpoint synthesis. The motivation, that is, the need for spectral radiance fields is: spectral information can provide more details on the material constitution of objects in the scene, which have been utilized in classic vision tasks, such as material classification (Jumanazarov et al. 2022). The idea of importing spectral information to rendering is a new perspective, which may provide inspiration for rendering and vision tasks.

Theoretically, we modify the traditional spectral rendering pipeline into two steps. According to the variant rendering pipeline, we first design an MLP-based architecture, which maps from an input 5D coordinate (3D position and 2D viewing direction) of real and synthetic scenes to properties of the scene (volume density and spectral radiance) at that location, named SpectralMLP. Volume rendering is applied to composite these values into discrete spectrum maps (Fig. 1 (a)). Then, SAUNet is proposed to combine these spectrum maps into high-quality RGB images (Fig. 1 (b)). In order to extract the spectral information better, we design the Spectrum Attention (SA) module to better explore the correlations between spectrum maps. The pipeline can be transferred to existing NeRF-based methods to promote their performance if we select their architecture as the baseline of SpectralMLP. We capture the real-world scenes and render the spectral datasets that contain spectrum maps and RGB images to optimize the outputs of SpectralMLP and the SAUNet, respectively.

Overall, our contributions can be summarized as:

- We propose SpectralNeRF, that builds up physically-based rendering with NeRF from the spectral perspective, which leads to mutual enhancement of spectral rendering and NeRF-based methods.
- We design the SAUNet to fuse the discrete spectrum maps to generate high-quality RGB images, which can approach the integral calculation in spectral rendering.
- We render 8 spectral datasets and capture 2 real-world scenes with spectrum maps and RGB images, and provide comprehensive comparisons of these datasets with several NeRF-based methods to demonstrate the superiority of the SpectralNeRF.

Related Works

Neural Radiance Field for 3D Scenes. Using the neural network to represent a 3D scene and generate novel views with weights of MLP or other network parameters has become a hot topic. Previous methods address the issue with explicit discrete representations (Dai, Song, and Xin 2015; Aliev et al. 2020; Wu, Xia, and Wang 2020; Mildenhall et al. 2019; Waechter, Moehrle, and Goesele 2014). Since Mildenhall *et al.* introduced the differentiable volumetric rendering technique to optimize a neural radiance field (Mildenhall et al. 2020), a number of studies have been carried out to dive deeper into NeRF-based architecture, including more detail preservation methods (Barron et al. 2021, 2022; Chen et al. 2022; Dave, Zhao, and Veeraraghavan 2022), the faster training and inference of NeRF (Martin-Brualla et al. 2021; Reiser et al. 2021), the extension from image to video (Li et al. 2021, 2023b), the refractive novel-view synthesis (Bemana et al. 2022), the dynamic scenes (Cao and Johnson 2023), the LiDAR scenes (Huang et al. 2023), the infrared and spectral scenes (Poggi et al. 2022), and the event cameras (Rudnev et al. 2023). It is challenging for these methods to represent scenes with complex textures. To solve such problems, we present a novel NeRF-based architecture, which introduces spectral information into the radiance field to simplify complicated scenes.

Spectral Rendering. With the development of computing power, various rendering technologies are proposed to obtain photo-realistic images (Nguyen-Phuoc et al. 2018; Peters et al. 2019; Dai et al. 2020; Li et al. 2022, 2023a; Hu et al. 2023). Spectral rendering is a more physically correct technique that indeed models a scene's light transport with real wavelengths (Wilkie and Purgathofer 2002). Over the past decades, many physically-based spectral rendering methods have been proposed, including stochastic sampling over the visible light (Watanabe et al. 2013), representing spectral information using Fourier coefficients (Peters et al. 2019) and sampling in the spatial domain (Knaus and Zwicker 2011; Watanabe et al. 2013). These methods are designed for canonical ray-tracing rendering pipelines, which might be time-consuming when rendering scenes with complicated geometry. We consider the scene information along the ray and take advantage of NeRF-based architecture to combine the neural radiance field and spectral information to perform physically-based spectral rendering.

Preliminaries of Spectral Rendering

Given the CIE tristimulus values X, Y and Z, the CIE color matching functions $f_X(\lambda)$, $f_Y(\lambda)$ and $f_Z(\lambda)$ involve the influence of light with wavelength λ to the three values (Color and Labs 1995), which were defined by measuring the mean color perception of a sample of human observers over the visual range from $\lambda_{\text{violet}} = 380$ to $\lambda_{\text{red}} = 780$ nanometer (nm). The following equation calculates the CIE X, Y, and Z values for light with wavelength λ:

$$\begin{cases} X = \kappa \sum f_X(\lambda) L(\lambda) \Delta\lambda \\ Y = \kappa \sum f_Y(\lambda) L(\lambda) \Delta\lambda \\ Z = \kappa \sum f_Z(\lambda) L(\lambda) \Delta\lambda, \end{cases} \tag{1}$$

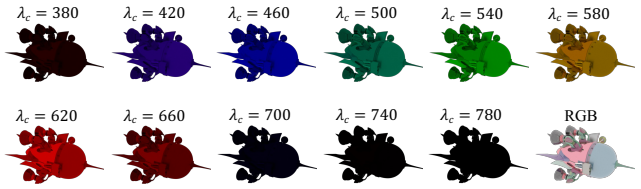


Figure 2: The RGB spectrum maps of different wavelengths and the RGB image of white-light illumination.

where κ is a normalizing constant, $\sum$ represents the summation of visible light, $\Delta\lambda$ represents the sampling interval, and L denotes the spectral power distribution of the light source from the direction $(\theta_{\mathrm{v}}, \varphi_{\mathrm{v}})$ of observation x:

$$L\left(x, \theta_{\mathrm{v}}, \varphi_{\mathrm{v}}, \lambda\right) = \int_{\Omega} f_{\mathrm{r}}\left(x, \theta, \varphi, \theta_{\mathrm{v}}, \varphi_{\mathrm{v}}, \lambda\right) R_{\mathrm{i}}(x, \theta, \varphi, \lambda) \cos\theta \, \mathrm{d}\omega, \tag{2}$$

where R_{i} represents the radiance from direction (θ, φ) to point x, Ω is the hemispherical space on the surface where point x is located, f_{r} represents the bidirectional reflectance distribution function (BRDF), which is determined by the reflection characteristics of the material at point x, $\mathrm{d}\omega$ is a solid angle. Note that, Eq. 1 is calculated with the form of summation to estimate the original continuous integral.

To obtain a colorimetrically correct RGB image, the X, Y, and Z values are transformed to the sRGB color space using:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.133 & -1.616 & -0.490 \\ -0.978 & 1.916 & 0.033 \\ 0.072 & -0.229 & 1.405 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \tag{3}$$

Due to inconsistency of application environments, there are various XYZ $\rightarrow$ RGB conversion methods (Smits 1999). The matrix M^{c} listed here is computationally convenient.

Obtaining L from the rendering machine is difficult. In order to simplify the rendering pipeline of $L \rightarrow$ XYZ $\rightarrow$ RGB, we embed the Eq. 3 to Eq. 1 to obtain the RGB spectrum maps corresponding to the wavelengths. Linearly combining the two equations to generate the RGB spectrum maps is reasonable on the one hand, and is simple yet effective on the other hand. The RGB spectrum maps corresponding to the wavelengths can be formulated as:

$$\begin{cases} R_\lambda = (M^{\mathrm{c}}_{11} f_X(\lambda) + M^{\mathrm{c}}_{12} f_Y(\lambda) + M^{\mathrm{c}}_{13} f_Z(\lambda)) L(\lambda) \Delta\lambda \\ G_\lambda = (M^{\mathrm{c}}_{21} f_X(\lambda) + M^{\mathrm{c}}_{22} f_Y(\lambda) + M^{\mathrm{c}}_{23} f_Z(\lambda)) L(\lambda) \Delta\lambda \\ B_\lambda = (M^{\mathrm{c}}_{31} f_X(\lambda) + M^{\mathrm{c}}_{32} f_Y(\lambda) + M^{\mathrm{c}}_{33} f_Z(\lambda)) L(\lambda) \Delta\lambda. \end{cases} \tag{4}$$

Finally, the RGB spectrum maps are combined to generate the RGB image of white-light illumination:

$$\begin{cases} R = \kappa \sum R_\lambda \\ G = \kappa \sum G_\lambda \\ B = \kappa \sum B_\lambda. \end{cases} \tag{5}$$

Figure 2 shows an example that includes 11 RGB spectrum maps and one RGB image rendered with our hypothesis by Mitsuba. The λ_c in Fig. 2 represents the center of the sampling interval of spectral illuminates.

Method

We propose an end-to-end NeRF-based architecture to achieve the physically-based spectral rendering from a novel perspective. As shown in Fig. 3, the architecture includes two modules operating the RGB spectrum map rendering (Fig. 3 (a)) and the spectrum fusion (Fig. 3 (c)). The first module is an MLP-based network to produce spectrum maps according to the given ray origin $\mathbf{o}$ and ray direction $\mathbf{d}$. The second module fuses the discrete spectrum maps to obtain an RGB image of white-light illumination, which applies the attention mechanism to better extract useful information from spectrum maps to approach the integral calculation in spectral rendering. Note that, we generate s_{num} discrete spectrum maps by uniform sampling through the visible light $380nm$–$780nm$. The overall color of the generated spectrum maps conforms to the distribution of the CIE color matching function in Fig. 3 (b). Both spectral rendering and NeRF-based methods will be improved through the proposed pipeline.

Spectral Radiance Field

We represent the scene as spectral radiance fields within bounded 3D volumes. For a given ray origin $\mathbf{o} = (x, y, z)$ and ray direction $\mathbf{d} = (\theta, \phi)$, we propose the SpectralMLP $F_{\mathbf{\Theta}}$ to generate the spectral radiance s_{λ_i} and the density σ of the ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$. SpectralMLP $F_{\mathbf{\Theta}}$ achieves the mapping from $(\mathbf{o}, \mathbf{d})$ to (s_{λ_i}, σ), which is defined as:

$$\left(\mathbf{s}_{\lambda_i}, \sigma\right) = F_{\mathbf{\Theta}}\left(\gamma(\mathbf{o}), \gamma(\mathbf{d})\right), \tag{6}$$

where $i \in \{1, 2, ..., s_{\text{num}}\}$, γ represents the positional encoding (Rahaman et al. 2019) that maps the inputs into higher dimensional frequency space, which is applied separately to each of the three coordinate values in $\mathbf{x}$ and to the three components of the direction unit vector $\mathbf{d}$. The s_{num} is set to 11 to achieve the balance between performance and efficiency.

The SpectralMLP finally outputs s_{num} spectral radiance and one σ value. As for different wavelength λ, the density of each point is same, but the spectral radiance is different. Volume rendering (Levoy 1990) is then applied to render the spectral radiance s_{λ_i} of each ray passing through the scene. The spectrum value $\widehat{\mathbf{S}}_{\lambda_i}(\mathbf{r})$ of ray $\mathbf{r}(t)$ is computed as:

$$\widehat{\mathbf{S}}_{\lambda_i}(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{s}_{\lambda_i}(\mathbf{r}(t), \mathbf{d})dt, \tag{7}$$

where $T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\mathbf{r}(p))dp\right)$, t denotes a position along the ray, t_n and t_f are the near and far boundary.

SAUNet

Theoretically, according to Eq. 5, directly combining the spectrum maps of every wavelength among the visible light is practicable. Rendering each spectral dataset with around 400 images may be adequate for the linear combination. However, in our implementation, such an operation is insufficient because the spectral datasets are extremely sparse. Therefore, we propose the Spectrum Attention UNet (SAUNet) (Fig. 4 (a)) to learn the correlations of spectrum maps and generate high-quality RGB outputs. Applying the SAUNet can imitate the original integral operation better

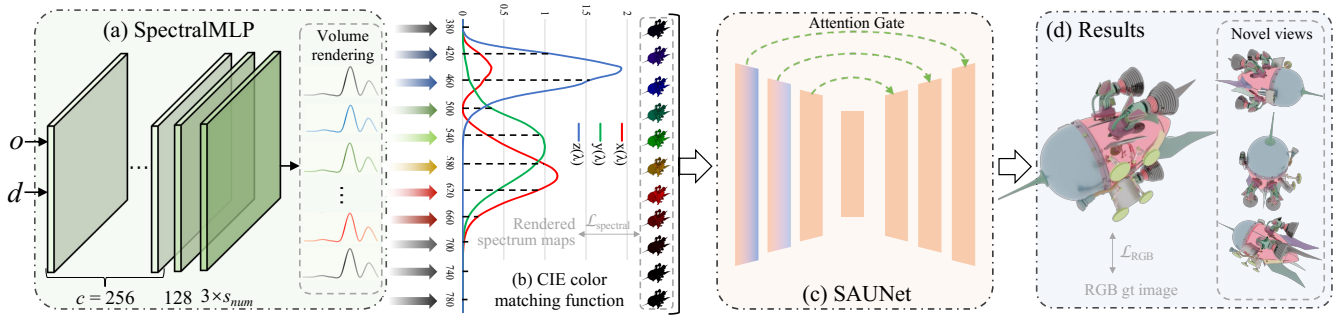


Figure 3: An overview of the SpectralNeRF. We first design (a) SpectralMLP to construct the spectral radiance field and generate s_{num} RGB spectrum maps with novel views using volume rendering. The generated spectrum maps are constrained by the rendered spectral images. The color of the spectrum maps matches the distribution of the CIE color matching function in (b). The (c) SAUNet combines the discrete spectrum maps to produce high-quality RGB outputs, constrained by the rendered RGB images. $\mathbf{o}$ is the ray origin, $\mathbf{d}$ is the ray direction and c represents the channels of different layers in SpectralMLP.

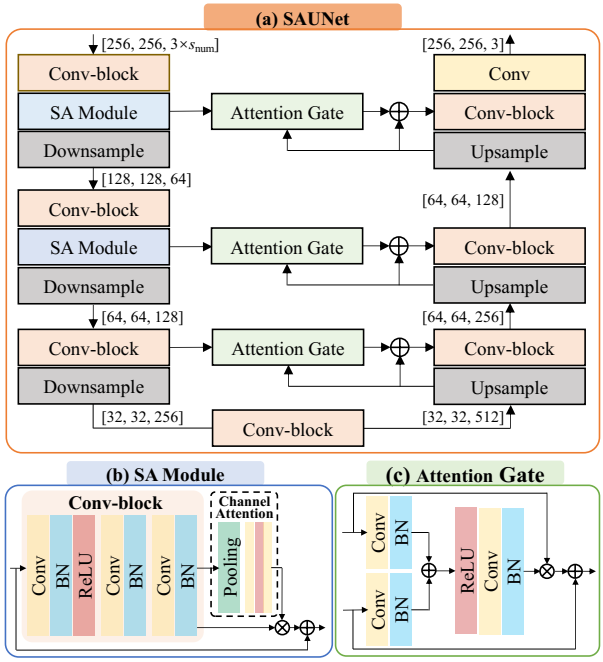


Figure 4: The detailed architecture of SAUNet.

and produce results close to the ground truth. We first introduce the Attention Gate (Oktay et al. 2018) (Fig. 4 (c)) to refine features from the encoder. The high-level features before outputting to the next decoder stage are guided by the low-level features from the encoder using the attention mechanism. We further design a Spectrum Attention (SA) module to better explore the correlations of spectrum maps.

The standard residual convolutional network is insufficient for extracting the spectral dependencies (Jiang et al. 2022). We modify the standard residual blocks and introduce the SA module (Fig. 4 (b)) to better combine the discrete spectrum maps to make the results close to the integral calculation in spectral rendering. Specifically, three 1×1 convolutional blocks are first used to reorganize and reweight the importance of spectrum maps. The channel attention (CA) (Hu, Shen, and Sun 2018) is then introduced to focus on inter-spectral feature fusion by attention mechanism in the channel dimension.

SAUNet contains 3 encoders and 3 decoders. Skip connections with Attention Gate pass feature maps from each encoder to decoder. Feature maps from low-levels contain more detailed information of spectrum maps. The SA module is placed in the first two encoders to best use the spectral information because higher-level features are more abstract, which may affect the ability of the network to explore the correlations. The fusion process is defined as:

$$\widehat{C} = \text{SAUNet}(\widehat{\mathbf{S}}_{\lambda_i}), \tag{8}$$

where $\widehat{C}$ represents the final output RGB image.

Optimization

The SpectralMLP involves spectral color transformations and the SAUNet imitates integral operations for spectrum maps. The objective function includes the following two items: 1) the weighted spectrum map reconstruction loss $\mathcal{L}_{\text{spectral}}$, which pushes the SpectralMLP to produce desired spectrum maps; 2) the RGB reconstruction loss $\mathcal{L}_{\text{RGB}}$, which optimizes the SAUNet to generate high-quality RGB images. The full objective function is described as:

$$\mathcal{L} = \mathcal{L}_{\text{spectral}} + \lambda_{\text{RGB}}\mathcal{L}_{\text{RGB}}, \tag{9}$$

where λ_{RGB} is a hyper-parameter to balance the contributions of the two losses. We empirically set it to 1.1.

Weighted Spectrum Map Reconstruction Loss. We found the image quality of generated spectrum maps is different, with its power concentrated in wavelength near $380nm$ and $780nm$ tends to be black, resulting in higher PSNR scores. Therefore, we propose the weighted spectral reconstruction loss $\mathcal{L}_{\text{spectral}}$ to better acquire useful information from more informative spectrum maps distributed in the middle of visible light. Similar to NeRF, we simultaneously optimize a coarse model and a fine model, and the loss is defined as:

$$\begin{aligned}\mathcal{L}_{\text{spectral}} = \textstyle\sum_i^{s_{\text{num}}} w_s \cdot \sum_{\mathbf{r}\in\mathcal{R}(\mathbf{P})} (||\widehat{\mathbf{S}}^c_{\lambda_i}(\mathbf{r}) - \mathbf{S}_{\lambda_i}(\mathbf{r})||_2^2 + \\ ||\widehat{\mathbf{S}}^f_{\lambda_i}(\mathbf{r}) - \mathbf{S}_{\lambda_i}(\mathbf{r})||_2^2),\end{aligned} \tag{10}$$

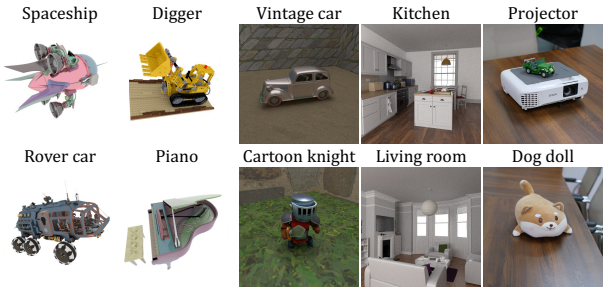


Figure 5: Scenes used for our synthetic and real datasets.

where $\mathbf{S}_{\lambda_i}$ represents the spectrum maps, $\mathcal{R}(\mathbf{P})$ is a set of camera rays at target position $\mathbf{P}$, $\widehat{\mathbf{S}}^c$ and $\widehat{\mathbf{S}}^f$ represent the spectrum maps generated in the coarse stage and fine stage, and w_s are the weights that are correlated to the PSNR scores of spectrum maps:

$$w_s = 2^{P_{\max}/P_\lambda}, \tag{11}$$

where $P_{\max}$ is the maximum PSNR value of s_{num} spectrum maps, and P_λ is the PSNR value of wavelength λ.

RGB Reconstruction Loss. The RGB reconstruction loss $\mathcal{L}_{\text{RGB}}$ minimizes the difference between the predicted RGB image $\widehat{C}$ and the rendered RGB ground truth C. The $\mathcal{L}_2$ distance is adopted as the loss function, written as:

$$\mathcal{L}_{\text{RGB}} = ||\widehat{C} - C||_2^2. \tag{12}$$

Datasets and Implementations

Datasets

Synthetic Scenes. We first render 6 scenes with models located in the middle of the field, among which 4 scenes are designed without background and the viewpoints are captured on a sphere surrounding the models (the first two columns of Fig. 5). The other 2 scenes use texture to construct the walls and floors, and their viewpoints are sampled on the upper hemisphere (the third column). Then, we render 2 indoor forward-facing scenes with limited camera location and camera perspectives (the fourth column).

Real-world Scenes. We capture 2 forward-facing real datasets in a sealed room (the last column of Fig. 5) using a camera and 8 color absorbers whose center wavelengths range from $400nm$ to $750nm$ with the interval of $50nm$. Different color absorbers are covered to the camera lens to obtain the spectral images. Each real-world dataset includes approximately 40 viewpoints.

Implementation Details

We implement the SpectralMLP on top of NeRF (Mildenhall et al. 2020), which uses an eight-layer MLP with 256 channels and ReLU activation to predict the density σ, and following two fully-connected layers with 128 and $3 \times s_{\text{num}}$ channels to obtain the spectral radiance. We sample 64 points along each ray in the coarse model and 128 points in the fine model on the dataset. Adam optimizer (Kingma and Ba 2015) is used for the SpectralMLP and the SAUNet, and their learning rate is set to 5×10^{-4} and 0.001, respectively.

Experiments

Quantitative Comparisons

We report quantitative performance using PSNR (higher is better), SSIM (higher is better) and LPIPS (Zhang et al. 2018) (lower is better). Table 1 shows the results on synthetic datasets, among which the top part is the results on 4 scenes without background, and the bottom part displays the results on 2 scenes using texture as background and 2 forward-facing scenes. Table 2 shows the results on 2 real-world datasets. Note that, Mip-NeRF is not applicable to forward-facing cases. As for challenging real-world scenes, the SAUNet produces slightly blurry results while preserving satisfactory image structure and contents, which leads to slightly lower LPIPS scores. This is a common trade-off in image restoration tasks. Benefiting from the spectral radiance field, our method outperforms other methods because it simplifies complicated scenes. On average, our method achieves 5% improvements compared to other methods.

Qualitative Comparisons

We present the qualitative comparisons in Fig. 6, Fig. 7 and Fig. 8. NeRF may cause aliasing when rendering views of varying resolutions. Mip-NeRF extends NeRF to instead reason about volumetric frustums along a cone, but fails to recover the detailed geometry and appearance when handling difficult cases. Aug-NeRF uses worst-case perturbations to regularize the model. However, the operation lacks robustness, and sometimes brings negative effects. Results in Fig. 6 clearly demonstrate that our method generates new viewpoints which are the closest to the ground truth. Other methods can reconstruct the low-frequency geometry, but fail to generate high-quality fine details. The results of NeRF and Aug-NeRF in Fig. 7 and Fig. 8 also cannot reconstruct fine details of the scene and generate results with fog artifacts. By concentrating on different spectral components of the complicated scene, our method outperforms the comparison methods by preserving more details. We then exhibit several spectrum maps of the SpectralMLP in Fig. 9.

Computational Times

The comparisons of inference times are reported in Table 3. Mip-NeRF and Aug-NeRF design complicated technologies to improve their performance, and therefore costing more time. Some recent methods focus on efficient training or inference. Nevertheless, it is challenging to achieve superiority in both speed and quality. The speed of our method is comparable to general NeRF methods. Our method is a little slower than NeRF. The volume rendering for s_{num} spectrum maps takes most of the extra time.

Ablation Studies

Effectiveness of Main Components. We conduct ablations on several components to understand how these main modules work, including s_{num}, Attention Gate, SA module, and the weights w_s in Eq. 10. The results are shown in Table 4. First, as shown in Table 4 (a) and (b), introducing the spectral radiance fields can effectively improve the performance.

	Spaceship			Rover car			Digger			Piano		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF	30.126	0.9358	0.0275	27.350	0.8920	0.0618	30.658	0.9187	0.0413	31.667	0.9239	**0.0394**
Mip-NeRF	31.495	0.9475	0.0535	30.028	0.9210	0.0376	33.301	0.9290	0.0435	31.872	**0.9304**	0.0630
Aug-NeRF	30.929	0.9402	0.0389	27.275	0.9022	0.0512	31.538	0.9248	0.0341	31.876	0.9229	0.0471
Ours	**31.951**	**0.9482**	**0.0250**	**30.086**	**0.9212**	**0.0356**	**33.378**	**0.9357**	**0.0259**	**32.266**	0.9290	0.0411

	Vintage car			Cartoon knight			Kitchen			Living room		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF	33.478	0.7958	**0.1319**	34.485	0.9273	0.1545	34.583	0.8943	0.1650	33.172	0.9929	0.0578
Mip-NeRF	33.883	0.8166	0.1747	**35.102**	0.9572	0.1526	–	–	–	–	–	–
Aug-NeRF	33.639	0.8002	0.1536	33.908	0.9287	0.1705	34.480	0.9026	0.1603	32.205	0.9649	0.0706
Ours	**34.480**	**0.8169**	0.1499	34.915	**0.9573**	**0.1510**	**35.115**	**0.9117**	**0.1637**	**33.665**	**0.9931**	**0.0479**

Table 1: Quantitative comparisons with other NeRF-based methods in terms of PSNR, SSIM and LPIPS on 8 synthetic datasets. The best and second-best results are marked in **bold** and underlined for better comparison.

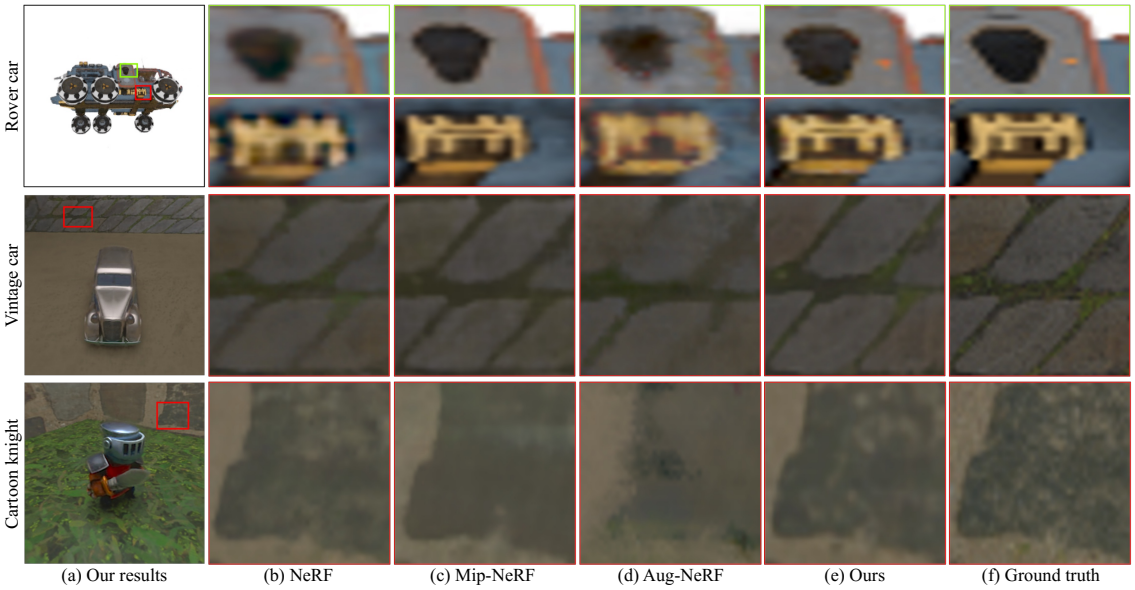


Figure 6: Qualitative comparisons with recent NeRF-based methods. In the first row, the comparison methods cannot recover the orange point, among which NeRF blurs it, Mip-NeRF and Aug-NeRF smooth it. The SpectralNeRF preserves more details.

	Projector			Dog doll		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF	28.967	0.943	0.047	22.504	0.874	0.132
Aug-NeRF	30.080	**0.957**	**0.035**	21.207	0.892	**0.117**
Ours	**31.254**	0.944	0.061	**25.126**	**0.893**	0.145

Table 2: Quantitative comparisons on 2 real-world scenes.

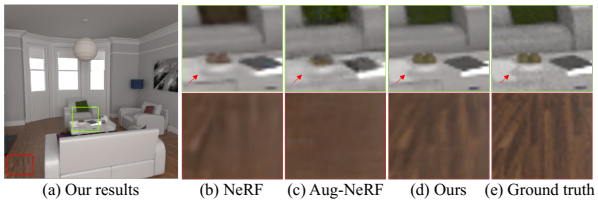
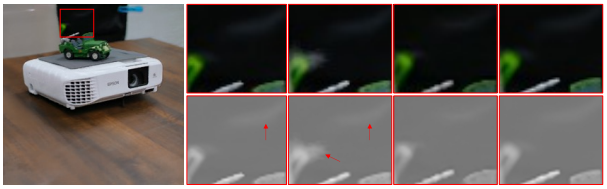


Figure 7: Comparisons on the synthetic forward-facing scene. NeRF and Aug-NeRF cannot reconstruct the shading of the fruit platter and the texture of the floor.

Second, as shown in Table 4 (g) and (h), removing the Attention Gate (AG) will degrade the results. Third, as shown in Table 4 (f) and (h), removing the weights w_s in Eq. 10 also affects the performance. Fourth, Table 4 (c), (d), (e), and (h) show the results when embedding the SA module to different encoder blocks. The SA module is placed in the first two encoders to best explore the correlations of spectrum maps.

The Role of SAUNet. We conduct the ablation study when replacing the SAUNet network with a simple full-connect (FC) layer to demonstrate the importance and effectiveness of the SAUNet. Table 5 shows corresponding results of several datasets, and the performance of the FC layer is



(a) Our results (b) NeRF (c) Aug-NeRF (d) Ours (e) Ground truth

Figure 8: Comparisons on the real-world scene. NeRF and Aug-NeRF get results with fog artifacts. We modify the contrast and the color to make the comparisons clear.

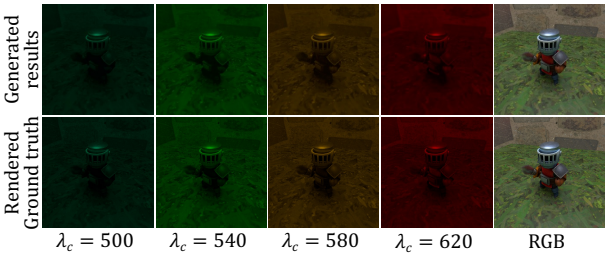


Figure 9: Several generated spectrum maps.

	NeRF	Mip-NeRF	Aug-NeRF	Ours
Time (s)	2.714	14.054	8.366	2.852

Table 3: Inference time on testsets with resolution 256×256.

	s_{num}	AG	SA	w_s	PSNR↑	SSIM↑	LPIPS↓
(a)	0			✓	30.126	0.9358	0.0275
(b)	11			✓	31.375	0.9429	0.0285
(c)	11	✓		✓	31.507	0.9445	0.0261
(d)	11	✓	E1	✓	31.552	0.9447	0.0266
(e)	11	✓	E1-E3	✓	31.723	0.9446	0.0263
(f)	11	✓	E1, E2		31.601	0.9449	0.0268
(g)	11		E1, E2	✓	31.875	0.9479	**0.0244**
(h)	11	✓	E1, E2	✓	**31.951**	**0.9482**	0.0250

Table 4: Ablation studies of different components. AG is the Attention Gate. SA is the Spectrum Attention module. E1, E2, and E3 are the first, second, and third encoder blocks. w_s represents the weights in Eq. 10.

obviously inferior to the SAUNet.

Ablation Study of MLP. The SpectralNeRF can be considered as an improvement technique of existing methods, which can promote their performance if we apply their network architecture as the baseline of SpectralMLP. We conduct the experiments when transferring the spectral radiance fields and SAUNet to Aug-NeRF (named S-Aug-NeRF), and the improvements of PSNR scores are listed in Table 6.

The Number of Spectral Maps. Figure 10 (Left) shows the PSNR and SSIM scores when setting the number of spectrum maps to 0 (vanilla NeRF), 6, 11, and 21. Obviously, reducing the output dimension ($11 \rightarrow 6$) slightly affects the performance, while the effect of changing from 11 to 21 is minor. To achieve the balance, we set the number of spectrum maps to 11 for the synthetic datasets and 8 for the real-world datasets. Second, Fig. 10 (Left) shows the results when setting the outputs of the SpectralMLP as three RGB channels, where the SAUNet is considered as a refinement network. That is, the spectral loss is removed. The results (3-dimension) are superior to the vanilla NeRF, while inferior to results with spectral information.

	Rover car	Digger	Vintage car	Mean
FC	27.309	30.617	33.382	31.430
SAUNet	30.086	33.378	34.480	33.648

Table 5: The PSNR scores for the FC and the SAUNet.

	Digger	Vintage car	Projector	Mean
Aug-NeRF	31.379	33.639	30.080	31.157
S-Aug-NeRF	31.506	34.130	32.385	32.636
Improvements	+0.127	+0.491	+2.305	+0.974

Table 6: The improvements of PSNR when selecting Aug-NeRF as the baseline of our SpectralMLP.

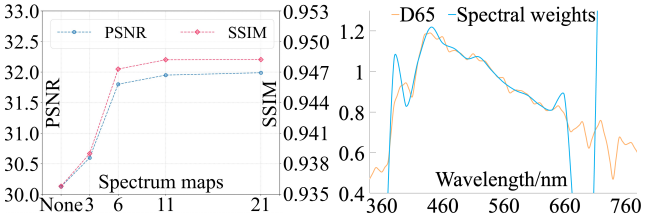


Figure 10: Left: The results when selecting different number of spectrum maps. Right: The correlations between CIE standard illuminant D65 and spectral weights.

The Variant Spectral Rendering Pipeline. To verify the correctness of the variant spectral rendering pipeline, we apply the least square (LSQ) method to obtain the adapted weights for spectrum maps. The estimated weights should be approximately proportional to the spectral power distribution $L(\lambda)$ of CIE standard illuminant D65 except for bands near $380nm$ and $780nm$ which are less informative. The almost identical curves in Fig. 10 (Right) demonstrate the rationality of the proposed variant spectral rendering.

Conclusion

We have proposed SpectralNeRF, an end-to-end NeRF-based method to achieve physically-based spectral rendering. We modified the traditional spectral rendering pipeline into two steps and designed SpectralMLP and SAUNet to build up the two steps. With the help of the spectral radiance field, our method can generate high-quality RGB output of white-light illumination. Comprehensive experiments have demonstrated the superiority of the proposed SpectralNeRF. In the future, we plan to reduce the number of spectrum maps, and construct each view with only one or sparse discrete spectral maps, which will broaden the applications.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (NSFC) under Grants No.62071097, No.62372091 and No.62031009, in part by Sichuan Science Foundation under Grants No.2023NSFSC0458 and No.2023NSFSC0462.

References

Aliev, K.-A.; Sevastopolsky, A.; Kolos, M.; Ulyanov, D.; and Lempitsky, V. 2020. Neural Point-based Graphics. In *Proc. ECCV*, 696–712.

Barron, J. T.; Mildenhall, B.; Tancik, M.; Hedman, P.; Martin-Brualla, R.; and Srinivasan, P. P. 2021. Mip-NeRF: A Multiscale Representation for Anti-aliasing Neural Radiance Fields. In *Proc. ICCV*, 5855–5864.

Barron, J. T.; Mildenhall, B.; Verbin, D.; Srinivasan, P. P.; and Hedman, P. 2022. Mip-NeRF 360: Unbounded Anti-aliased Neural Radiance Fields. In *Proc. CVPR*, 5470–5479.

Bemana, M.; Myszkowski, K.; Revall Frisvad, J.; Seidel, H.-P.; and Ritschel, T. 2022. Eikonal Fields for Refractive Novel-view Synthesis. In *Proc. ACM SIGGRAPH*, 1–9.

Bertrand, L.; Thoury, M.; Gueriau, P.; Anheim, É.; and Cohen, S. 2021. Deciphering the Chemistry of Cultural Heritage: Targeting Material Properties by Coupling Spectral Imaging with Image Analysis. *Accounts of Chemical Research*, 54(13): 2823–2832.

Cao, A.; and Johnson, J. 2023. HexPlane: A Fast Representation for Dynamic Scenes. In *Proc. CVPR*, 130–141.

Chen, T.; Wang, P.; Fan, Z.; and Wang, Z. 2022. Aug-NeRF: Training Stronger Neural Radiance Fields With Triple-Level Physically-Grounded Augmentations. In *Proc. CVPR*, 15191–15202.

Color; and Labs, V. R. 1995. Colour Matching Functions. http://cvrl.ioo.ucl.ac.uk. Accessed: 2024-01-26.

Dai, P.; Li, Z.; Zhang, Y.; Liu, S.; and Zeng, B. 2020. PBR-Net: Imitating Physically Based Rendering Using Deep Neural Network. *IEEE Trans. on Image Processing*, 29: 5980–5992.

Dai, Q.; Song, Y.; and Xin, Y. 2015. Random-accessible Volume Data Compression with Regression Function. In *International Conference on Computer-Aided Design and Computer Graphics*, 137–142.

Dave, A.; Zhao, Y.; and Veeraraghavan, A. 2022. PANDORA: Polarization-aided Neural Decomposition of Radiance. In *Proc. ECCV*, 538–556.

Helffer, B. 2013. *Spectral Theory and Its Applications*. 139. Cambridge University Press.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-Excitation Networks. In *Proc. CVPR*, 7132–7141.

Hu, T.; Xu, X.; Liu, S.; and Jia, J. 2023. Point2Pix: Photo-Realistic Point Cloud Rendering via Neural Radiance Fields. In *Proc. CVPR*, 8349–8358.

Huang, S.; Gojcic, Z.; Wang, Z.; Williams, F.; Kasten, Y.; Fidler, S.; Schindler, K.; and Litany, O. 2023. Neural LiDAR Fields for Novel View Synthesis. *arXiv preprint arXiv:2305.01643*.

Jiang, J.; Wang, C.; Liu, X.; Jiang, K.; and Ma, J. 2022. From Less to More: Spectral Splitting and Aggregation Network for Hyperspectral Face Super-Resolution. In *Proc. CVPR*, 267–276.

Jumanazarov, D.; Koo, J.; Poulsen, H. F.; Olsen, U. L.; and Iovea, M. 2022. Significance of the Spectral Correction of Photon Counting Detector Response in Material Classification from Spectral X-ray CT. *Journal of Medical Imaging*, 9(3): 034504–034504.

Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In *Proc. ICLR*.

Knaus, C.; and Zwicker, M. 2011. Progressive Photon Mapping: A Probabilistic Approach. *ACM Trans. Graph*, 30(3): 1–13.

Levoy, M. 1990. Efficient Ray Tracing of Volume Data. *ACM Trans. Graph*, 9(3): 245–261.

Li, R.; Dai, P.; Liu, G.; Zhang, S.; Zeng, B.; and Liu, S. 2023a. PBR-GAN: Imitating Physically Based Rendering with Generative Adversarial Networks. *IEEE Trans. on Circuits and Systems for Video Technology*.

Li, Z.; Niklaus, S.; Snavely, N.; and Wang, O. 2021. Neural Scene Flow Fields for Space-time View Synthesis of Dynamic Scenes. In *Proc. CVPR*, 6498–6508.

Li, Z.; Wang, L.; Huang, X.; Pan, C.; and Yang, J. 2022. PhyIR: Physics-Based Inverse Rendering for Panoramic Indoor Images. In *Proc. CVPR*, 12713–12723.

Li, Z.; Wang, Q.; Cole, F.; Tucker, R.; and Snavely, N. 2023b. DynIBaR: Neural Dynamic Image-Based Rendering. In *Proc. CVPR*, 4273–4284.

Liang, J.; Zhou, J.; Tong, L.; Bai, X.; and Wang, B. 2018. Material Based Salient Object Detection from Hyperspectral Images. *Pattern Recognition*, 76: 476–490.

Martin-Brualla, R.; Radwan, N.; Sajjadi, M. S.; Barron, J. T.; Dosovitskiy, A.; and Duckworth, D. 2021. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *Proc. CVPR*, 7210–7219.

Mildenhall, B.; Srinivasan, P. P.; Ortiz-Cayon, R.; Kalantari, N. K.; Ramamoorthi, R.; Ng, R.; and Kar, A. 2019. Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines. *ACM Trans. Graph*, 38(4): 1–14.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Proc. ECCV*, 405–421.

Newton, I. 1672. A Serie's of Quere's Propounded by Mr. Isaac Newton, to be Determin'd by Experiments, Positively and Directly Concluding His New Theory of Light and Colours; and Here Recommended to the Industry of the Lovers of Experimental Philosophy, as they Were Generously Imparted to the Publisher in a Letter of the Said Mr. Newtons of July 8.1672. *Philosophical Transactions of the Royal Society of London*, 7(85): 5004–5007.

Nguyen-Phuoc, T. H.; Li, C.; Balaban, S.; and Yang, Y. 2018. RenderNet: A Deep Convolutional Network for Differentiable Rendering from 3D Shapes. In *Proc. NIPS*, 7902–7912.

Oktay, O.; Schlemper, J.; Folgoc, L. L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N. Y.; Kainz, B.; et al. 2018. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv preprint arXiv:1804.03999*.

Peercy, M. S. 1993. Linear Color Representations for Full Speed Spectral Rendering. In *Annual Conference on Computer Graphics and Interactive Techniques*, 191–198.

Peters, C.; Merzbach, S.; Hanika, J.; and Dachsbacher, C. 2019. Using Moments to Represent Bounded Signals for Spectral Rendering. *ACM Trans. Graph*, 38(4): 1–14.

Pickholtz, R.; Schilling, D.; and Milstein, L. 1982. Theory of Spread-spectrum Communications - A Tutorial. *IEEE Trans. Communications*, 30(5): 855–884.

Poggi, M.; Ramirez, P. Z.; Tosi, F.; Salti, S.; Mattoccia, S.; and Di Stefano, L. 2022. Cross-Spectral Neural Radiance Fields. In *International Conference on 3D Vision*, 606–616.

Rahaman, N.; Baratin, A.; Arpit, D.; Draxler, F.; Lin, M.; Hamprecht, F.; Bengio, Y.; and Courville, A. 2019. On the Spectral Bias of Neural Networks. In *Proc. ICML*, 5301–5310.

Reiser, C.; Peng, S.; Liao, Y.; and Geiger, A. 2021. KiloNeRF: Speeding Up Neural Radiance Fields with Thousands of Tiny MLPs. In *Proc. ICCV*, 14335–14345.

Rudnev, V.; Elgharib, M.; Theobalt, C.; and Golyanik, V. 2023. EventNeRF: Neural Radiance Fields from A Single Colour Event Camera. In *Proc. CVPR*, 4992–5002.

Smits, B. 1999. An RGB to Spectrum Conversion for Reflectances. *Journal of Graphics Tools*, 4(4): 11–22.

Sun, Y.; Fracchia, F. D.; Drew, M. S.; and Calvert, T. W. 2001. A Spectrally Based Framework for Realistic Image Synthesis. *The Visual Computer*, 17(7): 429–444.

Uzair, M.; Mahmood, A.; and Mian, A. 2015. Hyperspectral Face Recognition with Spatiospectral Information Fusion and PLS Regression. *IEEE Trans. on Image Processing*, 24(3): 1127–1137.

Waechter, M.; Moehrle, N.; and Goesele, M. 2014. Let There be Color! Large-scale Texturing of 3D Reconstructions. In *Proc. ECCV*, 836–850.

Watanabe, S.; Kanamori, S.; Ikeda, S.; Raytchev, B.; Tamaki, T.; and Kaneda, K. 2013. Performance Improvement of Physically Based Spectral Rendering using Stochastic Sampling. In *International Workshop on Computational Color Imaging*, 184–198.

Wilkie, K. D. A. C. A.; and Purgathofer, W. 2002. Tone Reproduction and Physically Based Spectral Rendering. In *Eurographics*.

Wu, D.; Xia, S.-T.; and Wang, Y. 2020. Adversarial Weight Perturbation Helps Robust Generalization. In *Proc. NIPS*, 2958–2969.

Yan, Z.; Li, C.; and Lee, G. H. 2023. NeRF-DS: Neural Radiance Fields for Dynamic Specular Objects. In *Proc. CVPR*, 8285–8295.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The Unreasonable Effectiveness of Deep Features as A Perceptual Metric. In *Proc. CVPR*, 586–595.