Decoupling Scattering: Pseudo-Label Guided NeRF for Scenes with Scattering Media

Mingyang Zhang[1], **Junkang Zhang**[1], **Faming Fang** [1,2] *, **Guixu Zhang** [1]

[1]School of Computer Science and Technology, East China Normal University
[2]Shanghai Key Laboratory of Multidimensional Information Processing
{myzhang, jkzhang}@stu.ecnu.edu.cn, {fmfang, gxzhang}@cs.ecnu.edu.cn

Abstract

Neural Radiance Fields (NeRF) has been widely used in computer vision and graphics, achieving impressive results in novel view synthesis and multi-view 3D reconstruction. However, despite its excellent performance under ideal conditions, NeRF struggles in challenging environments such as hazy, foggy, and underwater scenes, primarily due to the difficulty in decoupling objects from the scattering medium. To mitigate this limitation, we proposed a novel approach for NeRF in scenes with scattering media. Specifically, we leverage pseudo-labels during the early stage of training to guide NeRF in decoupling the densities of objects and the scattering medium, guiding the model toward a more appropriate search space. Furthermore, we introduce a Cyclical Progressive Dimensional Optimization Strategy (CPDOS) that focuses on optimizing a single or a few variables during specific periods. Experimental results demonstrate that our method can effectively simulate hazy and underwater scenes, accurately decouple the scattering medium from objects, estimate atmospheric parameters, and outperform existing methods in novel view synthesis and image restoration tasks.

Code — https://github.com/MMMazart/DecoNeRF

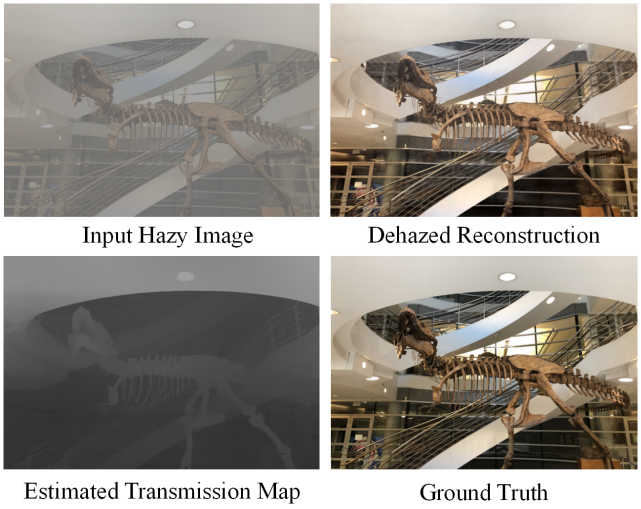

Figure 1: Our method can produce accurate dehazed images and transmission map estimation results.

Introduction

Neural Radiance Fields (NeRF) (Mildenhall et al. 2021) effectively reconstruct 3D scenes and generate novel views from multi-view 2D images using coordinate-based implicit representations, and has been widely used in computer vision and graphics. Although NeRF has demonstrated excellent performance under ideal conditions, it still struggles with challenging conditions, such as hazy, foggy and underwater scenes (Chen et al. 2023). The initial NeRF model assumes each ray path intersects only one opaque object and neglects scattering in the air (Levy et al. 2023), leading to its failure in handling environments with scattering media such as fog, haze and water, as it does not fully account for scattering effects.

When input images are severely degraded by scattering media like haze and water, NeRF reconstructions often exhibit blurred appearances and color distortions. Recent studies have proposed innovative approaches to address this challenge. One intuitive method treats the scattering medium as semi-transparent densities or voxels within the NeRF framework, extracting clear images by applying density thresholds, as seen in FogNeRF (Teigen et al. 2023) and (Jin et al. 2023). However, in practice, NeRF models haze as a coloration on object surfaces, complicating the distinction between objects and the scattering medium. Approaches like Seathru-NeRF (Levy et al. 2023), Dehaze-NeRF (Chen et al. 2023) and ScatterNeRF (Ramazzina et al. 2023) have been designed to handle scattering medium within rendering models, aiming to distinguish between medium and object densities using independent parameters within the NeRF framework. Despite these advances, it is highly under-constrained to decompose the scene into objects and medium components in an unsupervised way without robust constraint (Guo et al. 2022b). For instance, as illustrated in Figure 2, Seathru-NeRF (Levy et al. 2023) may converge to a trivial solution: one of the components explains the whole scene while leaving the other empty (Guo et al. 2022b).

Dehaze-NeRF (Chen et al. 2023) and ScatterNeRF (Ramazzina et al. 2023) incorporate handcrafted priors like

*Corresponding Author

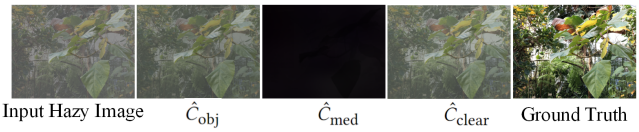

Figure 2: Training results of Seathru-NeRF: One component explained the entire scene, while the other remained nearly empty. (The meaning of $\hat{C}_{\text{obj}}$, $\hat{C}_{\text{med}}$ and $\hat{C}_{\text{clear}}$ is derived from Eq. 8 and Eq. 9)

Dark Channel Prior (DCP) (He, Sun, and Tang 2010) or Airlight Color Supervision (ACS) (Tang, Yang, and Wang 2014) as regularization terms to avoid trivial solutions. However, employing handcrafted priors as regularization terms offers limited effectiveness in constraining the training, as these empirically statistical priors may not be sufficient to describe complex real-world scenes, resulting in suboptimal dehazing performance (Wu et al. 2023). Water-NeRF (Sethuraman, Ramanagopal, and Skinner 2023) and WaterHE-NeRF (Zhou et al. 2023) use histogram equalization (HE) (Pizer et al. 1987) enhanced images as ground truth, optimizing through the Wasserstein distance minimization (Arjovsky, Chintala, and Bottou 2017), but this can lead to overfitting of the restoration results to the HE enhanced images.

Therefore, the key point in NeRF for reconstruction in scattering media is to find an effective way to decouple objects from the scattering medium. To address the issues of the methods mentioned above, we propose our method inspired by the concept of pseudo-labeling (Lee et al. 2013) in semi-supervised learning. In scenarios with scarce labeled data, pseudo-labels act as "soft supervision". Despite potential inaccuracies, they positively guide model training and enhance the extraction of information from unlabeled data. Pseudo-labels are commonly utilized in weakly-supervised and unsupervised learning. Therefore, in unsupervised tasks like reconstruction in scattering media using NeRF, we explore utilizing **pseudo-labels** during the early stage of training to guide the NeRF model in decoupling the densities of the scattering medium and the objects, thereby guiding the model toward a more reasonable search space. This approach effectively mitigates the limitations of handcrafted priors or regularization terms in some specific scenes, providing effective constraints to the NeRF training process. By introducing pseudo-labels only during the early stage of training, we prevent overfitting to the pseudo-labels.

However, more than this approach is needed to ensure a successful optimization. As Eq. 3, reconstructing the correct NeRF scene is a multi-dimensional optimization problem under degraded input conditions. We need to estimate the clear image J, the transmission map t, and the scattered light A, yet we rely solely on pseudo-labels to guide the reconstruction of clear image J, the solution space for the transmission map t and scattered light A remains vast. Inspired by the coordinate gradient descent method (Tseng and Yun 2009; Wright 2015), we propose the **Cyclical Progressive Dimensional Optimization Strategy (CPDOS).** This strategy focuses on optimizing a single or a few variables during specific periods. Through a cyclical and progressively refined optimization process, it carefully adjusts the restoration of images, the density of the scattering medium, and the atmospheric light, progressively narrowing down toward either the global optimum or a more favorable local optimum. Through such methodology, we can achieve more accurate restorations of images, transmission maps and estimates of atmospheric light, as illustrated in Figure 1.

Our contributions are summarized as follows:

- We introduce the use of **pseudo-labels** during the early stage of training, which guides the NeRF model toward a more reasonable search space by effectively decoupling object densities from the scattering medium. It mitigates the limitations of handcrafted priors or regularization terms in some specific scenes, providing constraints to the NeRF training process.
- We propose the **Cyclical Progressive Dimensional Optimization Strategy (CPDOS)**, which focuses on optimizing a single or a few variables during specific periods. This approach yields more accurate image restoration and novel view synthesis results while minimizing the reliance on guided images or priors.
- Experimental results substantiate the efficacy of our method, demonstrating its capability to successfully simulate scenes with scattering media, accurately decouple the scattering medium from objects, estimate atmospheric parameters, and achieve superior performance in image restoration and novel view synthesis tasks.

Related Work

Neural Radiance Fields (NeRF). As a novel method for view synthesis and 3D reconstruction, NeRF utilizes coordinate-based implicit 3D scene representation and differentiable volume rendering to achieve high-quality results in novel view synthesis and multi-view 3D reconstruction tasks. Some research aims to address the challenges of view synthesis and reconstruction under non-ideal image input conditions, enabling NeRF to adapt to more challenging conditions, such as blurry (Ma et al. 2022; Wang et al. 2023; Lee et al. 2023; Peng and Chellappa 2023), low-resolution (Wang et al. 2022a,b), low-light (Mildenhall et al. 2022; Cui et al. 2024) and reflective images (Verbin et al. 2022; Guo et al. 2022b; Zhu, Wan, and Shi 2022). These allow NeRF to adapt to more challenging environments.

Several recent works (Chen et al. 2023; Levy et al. 2023; Ramazzina et al. 2023; Li et al. 2023; Jin et al. 2023; Teigen et al. 2023; Zhang and Johnson-Roberson 2023; Sethuraman, Ramanagopal, and Skinner 2023; Zhou et al. 2023), including ours, have focused on addressing the reconstruction and image restoration challenges of NeRF models in scattering media such as haze and water. Methods like Seathru-NeRF (Levy et al. 2023), ScatterNeRF (Ramazzina et al. 2023), and DehazeNeRF (Chen et al. 2023) use rendering models that account for scattering effects. These approaches assign independent color and density parameters within the NeRF framework to decouple the densities of the medium and objects. ScatterNeRF (Ramazzina et al. 2023) employs physically inspired losses for scene reconstruction,

demonstrating its effectiveness on multi-view outdoor data and large-scale fog chamber datasets. Dehaze-NeRF (Chen et al. 2023) incorporates handcrafted physical priors like the DCP (He, Sun, and Tang 2010), as a regularization term to constrain model optimization. Unlike previous approaches, Dehazing-NeRF (Li et al. 2023) does not rely on handcrafted dehazing priors. Instead, it learns to model a clear scene by utilizing depth information of the clear scene and atmospheric parameters estimated by a pre-trained network to reconstruct hazy scenes. FogNeRF (Teigen et al. 2023) and (Jin et al. 2023) extract clear images from the radiance field by setting density thresholds. WaterNeRF (Sethuraman, Ramanagopal, and Skinner 2023) and WaterHE-NeRF (Zhou et al. 2023) use the results of HE (Pizer et al. 1987) as pseudo ground truth to supervise the optimization process of NeRF.

Image Restoration in Scattering Media. Image Restoration in scattering media like haze and water, is a critical task in computer vision and image processing. It aims to recover clear scenes from degraded images. These tasks are particularly challenging due to their ill-posed nature (He, Sun, and Tang 2010; Wu et al. 2023; Liu et al. 2019) and are typically approached using prior-based or learning-based methods.

Early work focused on leveraging image priors and physical models of light scattering to restore image clarity. For hazy images, the Dark Channel Prior (He, Sun, and Tang 2010) estimates the transmission map and atmospheric light by exploiting the low intensity of at least one color channel in clear regions. The Color Attenuation Prior (Zhu, Mai, and Shao 2015) uses brightness and saturation variations to distinguish hazy areas. Other effective priors include the haze-line theory (Fattal 2014), color ellipsoid prior (Bui and Kim 2017) and gamma correction prior (Ju et al. 2019). In the underwater domain, methods similar to DCP have been adapted for underwater image modeling(Yu et al. 2020; Liang et al. 2021), and approaches like the retinex-based method (Fu et al. 2014) have been proposed for enhancing single underwater images. The rank-one prior (ROP) (Liu et al. 2021), applicable to both hazy and underwater scenes, approximates the transmission map as a rank-one matrix for real-time recovery.

While prior-based methods are computationally efficient, they often struggle with complex real-world scenes, particularly when these scenes deviate from prior assumptions, leading to suboptimal results. For instance, DCP performs poorly in areas with large white regions, such as the sky. The advent of deep learning and large-scale datasets has significantly advanced image enhancement in scattering media. In dehazing, learning-based approaches like DeHamer (Guo et al. 2022a) combine the strengths of CNNs and Transformers, integrating global and local attention features for improved dehazing. C2PNet (Zheng et al. 2023) reinterprets the physical model by introducing a physically aware dual-branch unit, while RIDCP (Wu et al. 2023) incorporates codebook priors from high-quality images into the dehazing network for enhanced results. Similarly, in underwater image enhancement, learning-based methods have achieved substantial improvements. For instance, (Li et al. 2019) introduced an underwater image enhancement benchmark and network. WF-Diff (Zhao et al. 2024) utilizes wavelet-based Fourier interaction and frequency diffusion adjustment to enhance underwater images, effectively improving color accuracy and contrast for superior visual quality.

Method

Preliminary

Neural Radiance Fields. The original NeRF (Mildenhall et al. 2021) implicitly represents a differentiable 3D scene by a continuous function, which takes 3D position $\mathbf{x} = (x, y, z)$ and observed viewing direction $\mathbf{d} = (\theta, \phi)$ as input, and outputs the color $\mathbf{c}(\mathbf{x}, \mathbf{d}) = (r, g, b)$ and volume density $\sigma(\mathbf{x})$. NeRF is typically parametrized by a multilayer perceptron ($\mathbf{MLP}$) $f_\Theta : (\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma)$.

NeRF only focuses on the emission part of the classical volume rendering, the color along a camera ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$, coming from the camera's center $\mathbf{o}$ in the direction $\mathbf{d}$ can be calculated as:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})\mathrm{d}t, \tag{1}$$

where $T(t) = \exp\left(-\int_{t_n}^{t_f} \sigma(\mathbf{r}(t))\mathrm{d}t\right)$ is the accumulated transmittance between the ray section t_n to t_f, which means the probability that the ray travels from t_n to t_f without hitting other particles along the way, obtained by integrating the volume density $\sigma(\mathbf{r}(t))$.

In practice, to render a pixel, we sample N points from the camera ray $\mathbf{r} = \mathbf{o} + t\mathbf{d}$, obtain their densities and colors using the $\mathbf{MLP}$, and estimate the color $C(\mathbf{r})$ using volume rendering approximated by the quadrature rule (Max 1995):

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i \alpha_i \mathbf{c}_i, \tag{2}$$

where $\alpha_i = 1 - e^{(-\sigma_i \delta_i)}$ represents the occupancy probability of a point, $\delta_i = t_{i+1} - t_i$ is the distance between adjacent samples, and $T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right)$. The NeRF model is optimized by minimizing the reconstruction loss: $\mathcal{L} = \|\hat{C}(\mathbf{r}) - C(\mathbf{r})\|_2$.

Image Formation in Scattering Media. The formulation of images affected by scattering media is commonly described by a physical scattering model (Wu et al. 2023; McCartney 1976):

$$I(x) = t(x)J(x) + (1 - t(x))A, \tag{3}$$

where $I(x)$ denotes the degraded image and $J(x)$ is its corresponding clear image. The variables A and $t(x)$ are the global atmospheric light and transmission map, respectively. The transmission map $t(x) = e^{-\beta d(x)}$ depends on scene depth $d(x)$ and medium density coefficient β.

Neural Radiance Scattering Model

Similar to recent works like SeaThruNeRF (Levy et al. 2023), DehazeNeRF (Chen et al. 2023), and ScatterNeRF (Ramazzina et al. 2023), we extend the NeRF model using the Radiative Transfer Equation (RTE) (Chandrasekhar

2013) to handle scenes with scattering media, beyond free space with no scattering. To accommodate this expansion, we incorporated scattering medium into Eq. 1 and set separate color and density parameters for objects and the medium:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \left(\sigma^{\text{obj}}(t)\mathbf{c}^{\text{obj}}(t) + \sigma^{\text{med}}(t)\mathbf{c}^{\text{med}}(t)\right) dt, \tag{4}$$

where $T(t) = \exp\left(-\int_{t_n}^{t_f} \left(\sigma^{\text{obj}}(s) + \sigma^{\text{med}}(s)\right) ds\right)$. In practice, converted to discrete form, we get:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i^{\text{obj}} T_i^{\text{med}} \left(\alpha_i^{\text{obj}}\mathbf{c}_i^{\text{obj}} + \alpha_i^{\text{med}}\mathbf{c}_i^{\text{med}}\right). \tag{5}$$

The calculations of T_i and α_i are similar to those in Eq. 2.

Splitting the discretized color rendering equations into the "object" and "medium" components, we get:

$$\hat{C}_{\text{obj}}(\mathbf{r}) = \sum_{i=1}^{N} T_i^{\text{med}} T_i^{\text{obj}} \alpha_i^{\text{obj}} \mathbf{c}_i^{\text{obj}}, \tag{6}$$

$$\hat{C}_{\text{med}}(\mathbf{r}) = \sum_{i=1}^{N} T_i^{\text{obj}} T_i^{\text{med}} \alpha_i^{\text{med}} \mathbf{c}_i^{\text{med}}, \tag{7}$$

$$\hat{C} = \hat{C}_{\text{obj}} + \hat{C}_{\text{med}}. \tag{8}$$

To enhance model optimization efficiency and reduce complexity, we adopt a simplification strategy similar to Seathru-NeRF (Levy et al. 2023), constraining the medium parameters to remain constant along 3D viewing rays.

In clear scenes, σ^{med} is equal to zero, so reconstructing a clear scene only requires eliminating the scattering medium:

$$\hat{C}_{\text{clear}}(\mathbf{r}) = \sum_{i=1}^{N} T_i^{\text{obj}} \alpha_i^{\text{obj}} \mathbf{c}_i^{\text{obj}}. \tag{9}$$

Pseudo-Label Guided Density Decoupling

An intuitive approach involves using handcrafted priors or regularization terms to constrain NeRF model training, as seen with techniques like the DCP (He, Sun, and Tang 2010; Chen et al. 2023) and the ACS (Tang, Yang, and Wang 2014; Ramazzina et al. 2023). However, this may lead to over-constraints or failures in scenes that do not match the assumptions. Inspired by pseudo-labeling (Lee et al. 2013) in semi-supervised learning, we propose a more effective way to constrain the training process. During the early stage of training, pseudo-labels are applied to guide the NeRF model, decoupling object densities from the scattering medium and leading to a more reasonable search space. This approach mitigates the limitations of handcrafted priors or regularization terms. The specific procedure is as follows:

- **Generation of Pseudo-Labels:** We use existing dehazing techniques to create pseudo ground truth images representing the scene without the scattering medium, approximating object densities free from the influence of the scattering medium.

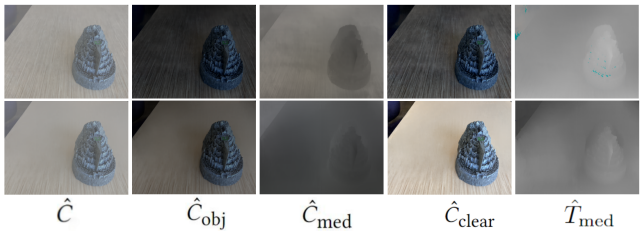

Figure 3: The first and second rows correspond to different sets of solutions. Higher transmission rates or brighter atmospheric scattering light result in darker restored image colors. $\hat{T}_{\text{med}}$ denotes the estimated transmission map.

- **Early Training with Pseudo-Labels:** Initially, we supervise training by minimizing the mean squared error (MSE) between the recovered images and pseudo-labels, helping the model decouple object densities from the scattering medium and guiding it toward a suitable search space. After a set number of iterations, we transition to standard training to prevent overfitting and allow the model to refine its understanding of the scene while considering scattering effects.

Cyclical Progressive Dimensional Optimization Strategy

Reconstructing NeRF scenes with scattering media requires estimating multiple parameters, as illustrated in Eq. 4. To simplify the explanation, we describe this process using a more straightforward formulation in Eq. 3. For the color of each pixel, we have $I = tJ + (1-t)A$, with only the ground truth value of I (hazy image pixel) available, we need to optimize t, J and A. Previous methods have predominantly relied on handcrafted priors (such as the DCP (He, Sun, and Tang 2010; Chen et al. 2023)) or regularization terms (such as the Clear Scene Entropy Minimization (Ramazzina et al. 2023)) to constrain J, with less focus on t and A, resulting in these parameters having a large solution space and multiple potential solutions. Therefore, merely constraining J while attempting to optimize all parameters might lead to a trival solution. As shown in Figure 3, there are many potential solutions that satisfy the conditions, higher transmission rates or brighter atmospheric scattering light will result in darker restored image colors. Therefore, accurate estimation of t and A is crucial for the precise recovery of J. We aim to support the recovery of J by obtaining more accurate estimates of t and A, which is a complex multidimensional optimization. Drawing on insights from the coordinate gradient descent method (Tseng and Yun 2009; Wright 2015), we propose the Cyclical Progressive Dimensional Optimization Strategy (CPDOS), which iteratively optimizes the clear image J, transmission map t, and atmospheric light A, as detailed in Algorithm 1[1]. By cycling through these dimensions,

[1]With pseudo-label supervision, the optimization path of J is largely predetermined, which allows us to treat J as "fixed". In phases 1 and 2, it may appear that we fix one variable while updating the other two, but since J's optimization path is mostly fixed,

Algorithm 1: Cyclical Optimization Phases for NeRF Model

Input: Hazy Image Pixel C, Pseudo-Label Pixel P
Parameter: J, t, A

1: **Initialization:** Random initialization of $J^{(0)}$, $t^{(0)}$, and $A^{(0)}$.
2: **Cyclical Optimization Phases:**
3: **for** each iteration cycle k **do**
4: **Phase 1:** Fixed $t^{(k)}$. Solve for $J^{(k+\frac{1}{3})}$ and $A^{(k+\frac{1}{3})}$ by minimizing:

$$\mathcal{L} = \|\hat{C} - C\|_2 + \|\hat{C}_{\text{clear}} - P\|_2.$$

5: **Phase 2: Fixed** $A^{(k+\frac{1}{3})}$. Solve for $J^{(k+\frac{2}{3})}$ and $t^{(k+\frac{2}{3})}$ by minimizing:

$$\mathcal{L} = \|\hat{C} - C\|_2 + \|\hat{C}_{\text{clear}} - P\|_2.$$

6: **Phase 3: Fixed** $t^{(k+\frac{2}{3})}$ **and** $A^{(k+\frac{2}{3})}$. Solve for $J^{(k+1)}$ by minimizing:

$$\mathcal{L} = \|\hat{C} - C\|_2.$$

7: **end for**
8: **Transition to Standard Training:** Transition to standard training to expedite NeRF model convergence.

CPDOS helps escape local optima, enabling the discovery of improved, three-dimensionally consistent solutions.

Implementation Details

Our implementation is based on the code released in Mip-NeRF-360 (Barron et al. 2022) and SeaThru-NeRF (Levy et al. 2023), they were implemented in Jax (Bradbury et al. 2018), while we re-implemented in PyTorch (Paszke et al. 2019). For the ScatterMLP, we utilize a linear layer with 256 features and a Softplus activation to predict σ^{med}, along with another linear layer with 256 features and a Sigmoid activation to predict c^{med}. We maintain the same learning rate and optimization parameters as used in Mip-NeRF-360 (Barron et al. 2022) and train each scene for 250k iterations on a single NVIDIA 3090 GPU, and we employ the Pseudo-Label Guided and Cyclical Progressive Dimensional Optimization Strategy during the first 50k iterations.

In practice, we utilized a codebook-based model called RQ-LLIE (Liu et al. 2023) to train a dehazing network. The pseudo-labels generated by this model served as guidance during the early training stage. Notably, this step can be substituted with other effective methods.

Experiments and Results

Datasets and Metrics

Hazy Scenes. Since capturing real-world multi-view hazy scenes is challenging, to evaluate the performance of our method, we synthesized hazy scene data using eight scenes from the well-known LLFF dataset. We utilized depth maps predicted by Mip-NeRF-360 (Barron et al. 2022) and applied the ASM (McCartney 1976) formula to haze the original clear images. The scattering factor β was randomly sampled from a uniform distribution in the range [0.7, 1.0], and the scattering light A was randomly sampled from a uniform distribution in the range [0.4, 0.6]. These parameters remained fixed across multiple images within a single scene to maintain three-dimensional consistency. Additionally, 10% of the images in each scene were set aside as a test set to evaluate the performance of novel view synthesis.

Real World Underwater Scenes. To validate the effectiveness of our method, we conducted qualitative experiments on two real-world underwater scenes from Seathru-NeRF (Ramazzina et al. 2023), including 20 images from the Red Sea and 18 from the Caribbean Sea, with three images in each set reserved for novel view synthesis evaluation.

Metrics. The quality of the rendered image is evaluated using the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), and Perceptual Similarity (LPIPS) metrics between the rendered image and the ground truth image.

Dehazing Scene Reconstruction

Given the limited research on reconstructing hazy scenes using NeRF in current open-source projects, we carefully selected some reasonable baseline methods for comparative analysis. Firstly, we chose Seathru-NeRF (Levy et al. 2023) as a reference point, an open-source method that holds a leading position in NeRF scene reconstruction within scattering medium. Additionally, we also selected several state-of-the-art single-image dehazing methods, and considering that they are typically not applicable for novel view synthesis or multi-view 3D reconstruction, we initially applied them to each hazy image to obtain the recovered images, which were then used as input for training the NeRF model. In the selection of dehazing methods, we selected two classical single-image dehazing methods: the famous Dark Channel prior (He, Sun, and Tang 2010) (DCP + NeRF) and the recent Rank One prior (Liu et al. 2021) (ROP + NeRF). Additionally, we chose two state-of-the-art learning-based single-image dehazing methods, namely DehazeFormer (Song et al. 2023) (DehazeFormer + NeRF) and MITNet (Shen et al. 2023) (MITNet + NeRF). We evaluated their results in novel view synthesis. We consistently used Mip-NeRF 360 (Barron et al. 2022) as the baseline method, referring to it simply as "NeRF" for convenience. And we denote the method that utilizes codebook (Liu et al. 2023) for pseudo-labels generation as "$ours$". Additionally, to ensure consistency and fairness in the experiments, all learning-based methods mentioned in this study were trained on the same dataset, with details provided in the supplementary materials.

Qualitative Evaluation. Figure 4 shows qualitative evaluation results of all methods on a subset of LLFF scenes. In some specific scenes, traditional handcrafted methods are prone to color distortions and error recovery results, such as "flower" and "leaves". This is often due to the mismatch between their assumptions and the actual scene conditions, as well as their inability to accurately estimate the atmospheric scattering parameters. Our proposed method over-

this can be approximated as fixing the paths of two variables and updating the third. This approach is consistent and symmetric.

Figure 4: Qualitative comparison of haze scene reconstruction using our method and other state-of-the-art dehazing methods. Our method demonstrates strong robustness, achieving better dehazing effects across various scenes than other methods.

	PSNR↑	SSIM↑	LPIPS ↓
DCP+NeRF	13.76	0.673	0.196
ROP+NeRF	17.53	0.724	0.198
Dehazemer+NeRF	18.76	0.724	0.194
MITNet+NeRF	19.92	0.727	0.215
Codebook+NeRF	17.85	0.697	0.265
SeaThru-NeRF	15.88	0.663	0.266
Ours	**21.42**	**0.740**	**0.166**

Table 1: Quantitative comparison of hazy scene reconstruction using our method and others. **Bold** indicates the best.

comes these limitations and provides more accurate restoration results. Although learning-based methods have shown a degree of generality, their dehazing effects are not significant when dealing with scenes such as "flower", "leaves" and "orchids". In contrast, our method can completely remove haze, presenting clearer visual results. Seathru-NeRF (Levy et al. 2023) shows limited dehazing effects in most scenes, with a tendency to incorrectly model haze as the surface color of objects. This issue stems from insufficient constraints during the model optimization process, leading to multiple solutions that satisfy the conditions and potentially converging to a trivial solution. Overall, our method demonstrates strong robustness, achieving satisfactory dehazing effects across various scenes.

Quantitative Evaluation. Table 1 presents the quantitative evaluation results of our method. For detailed experimental data, please refer to the extended version of the paper. Our method consistently ranked first or second in most scenes for all three key image quality metrics PSNR, SSIM, and LPIPS, achieving the highest average scores across all three metrics. This indicates superior pixel-level reconstruction quality, robust structural preservation, and high perceptual authenticity, confirming the method's effectiveness and robustness.

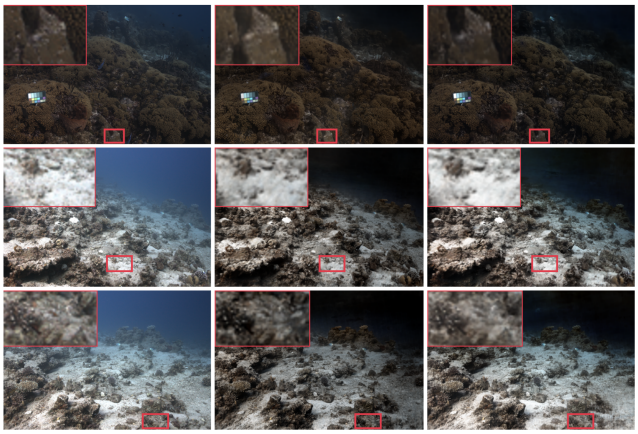

Figure 5: Reconstruction results for real-world underwater scenes. Row 1: Caribbean Sea; Rows 2 and 3: Red Sea.

Real World Underwater Scene Reconstruction

To validate the effectiveness of our method in real-world scenes, we conducted a qualitative comparison between our method without label guidance and Seathru-NeRF (Levy et al. 2023). Comparisons with other underwater enhancement methods are available within Seathru-NeRF. We present the qualitative results under novel view synthesis.

Qualitative Evaluation. Figure 5 presents the qualitative results of our method compared to Seathru-NeRF (Levy et al. 2023). Both methods demonstrate some degree of water removal in real-world underwater scenes. In Row 1, our method achieves better detail preservation, whereas Seathru-NeRF exhibits noticeable loss of detail. Rows 2 and 3 show that Seathru-NeRF tends to over-remove water in close-up places, leading to color distortion and dullness. In contrast, our method avoids these issues.

	PSNR↑	SSIM↑	LPIPS ↓	$\mathrm{MSE_T}$ ↓	$\mathrm{MSE_A}$ ↓
w/o CPDOS	19.49	0.699	0.188	0.0118	0.0069
w/o label guide	19.55	0.725	0.175	0.0132	0.0067
Ours	21.42	0.740	0.166	0.0036	0.0011

Table 2: Ablation Study. The proposed two components can improve the model's performance in various aspects.

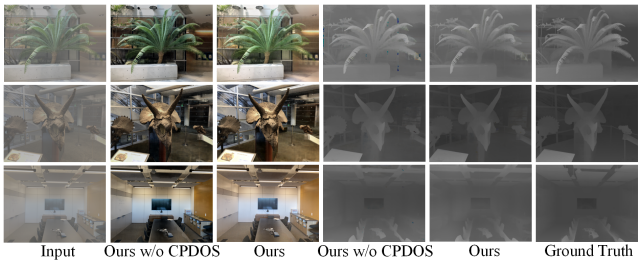

Figure 6: Examples of transmission map estimation by our method. We can achieve more accurate estimates of transmission maps compared to ours without CPDOS, resulting in better dehazing effects. The last three columns represent the transmission maps.

Ablation Study

Component Evaluation. To evaluate the role and contribution of the different components within our framework, we conducted an ablation study under the same experimental settings, the results presented in Table 2. We denote the MSE between the estimated transmission map and its ground truth as $\mathrm{MSE_T}$, and the MSE between the estimated atmospheric light and its ground truth as $\mathrm{MSE_A}$. The experimental results demonstrate that our components contributed to improvements across three image evaluation metrics and enhanced the accuracy of atmospheric parameters estimation, yielding better dehazing effects. As illustrated in Figure 6, Our method can estimate more accurate transmission maps, resulting in excellent dehazing effects.

Pseudo-Label Effectiveness Evaluation. To validate the effectiveness of pseudo-labels, we conducted experiments using two different pseudo-labeling methods: DehazeFormer and Codebook. As Table 3, our method achieved a PSNR of 21.74 with DehazeFormer (average PSNR: 19.11) and a PSNR of 22.23 with Codebook (average PSNR: 18.72), surpassing the quality of both pseudo-labels. Furthermore, our method also yields comparable or superior performance in SSIM and LPIPS metrics. For detailed experimental data, please refer to the extended version of the paper. As shown in Figure 4, our approach not only removed residual haze present in the pseudo-labels but also further enhanced the dehazing effect.

Insight into the Effectiveness of CPDOS. We conducted an additional experiment, revealing that using the ground truth of the clear image as a regularization term often leads to convergence at a local optimum. In contrast, incorporating CPDOS facilitates faster and more effective optimization, enabling the model to escape local optima and find a better solution among multiple local optima, as illustrated in Figure 7. Using pseudo-labels as a regularization term is simpler and more likely to ensure effectiveness.

	PSNR↑	SSIM↑	LPIPS ↓
DehazeFormer	19.11	0.762	0.173
DehazeFormer-guided	21.74	0.762	0.174
Codebook	18.72	0.748	0.223
Codebook-guided	22.23	0.778	0.145

Table 3: Pseudo-Label Effectiveness Evaluation

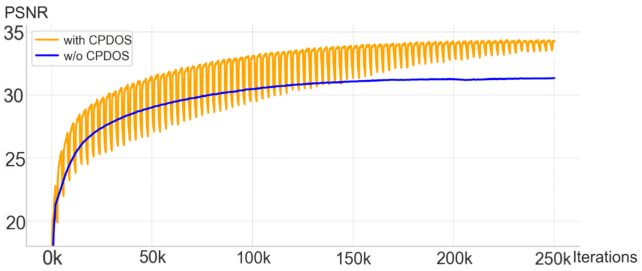

Figure 7: PSNR comparison over iterations: with CPDOS vs. without CPDOS.

Limitations and Conclusion

This paper introduces a novel method that effectively addresses the challenges NeRF faces in scenes influenced by scattering media. Our approach consistently achieves high-quality image restoration and view synthesis across various scenes, demonstrating significant robustness. Experimental results confirm that our method excels in simulating hazy and underwater environments, accurately decoupling the scattering medium from objects, estimating atmospheric parameters, and outperforming existing techniques in novel view synthesis and image restoration tasks.

However, our method has several limitations. Firstly, the performance of our method is inherently influenced by the accuracy of pseudo-labels, making it crucial to ensure that these pseudo-labels are good enough to guide the NeRF model toward a reasonable solution space. Developing an automated approach to assess the adequacy of pseudo-labels will be a focus of future research. Notably, our method has shown the ability to achieve superior results even with imperfect pseudo-labels, as confirmed by our experimental findings. If other deep learning methods perform well, then we can achieve even better results in NeRF, surpassing the performance of existing "deep learning + NeRF" approaches. Furthermore, this paper only covers some possible scenes, which will be a direction for future work.

Acknowledgements

This work was supported by the National Key R&D Program of China (2022ZD0161800), and the National Natural Science Foundation of China under Grant 62271203.

References

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein generative adversarial networks. In *International conference on machine learning*, 214–223. PMLR.

Barron, J. T.; Mildenhall, B.; Verbin, D.; Srinivasan, P. P.; and Hedman, P. 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5470–5479.

Bradbury, J.; Frostig, R.; Hawkins, P.; Johnson, M. J.; Leary, C.; Maclaurin, D.; Necula, G.; Paszke, A.; VanderPlas, J.; Wanderman-Milne, S.; and Zhang, Q. 2018. JAX: composable transformations of Python+NumPy programs.

Bui, T. M.; and Kim, W. 2017. Single image dehazing using color ellipsoid prior. *IEEE Transactions on Image Processing*, 27(2): 999–1009.

Chandrasekhar, S. 2013. *Radiative transfer*. Courier Corporation.

Chen, W.-T.; Yifan, W.; Kuo, S.-Y.; and Wetzstein, G. 2023. Dehazenerf: Multiple image haze removal and 3d shape reconstruction using neural radiance fields. *arXiv preprint arXiv:2303.11364*.

Cui, Z.; Gu, L.; Sun, X.; Ma, X.; Qiao, Y.; and Harada, T. 2024. Aleth-nerf: Illumination adaptive nerf with concealing field assumption. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 1435–1444.

Fattal, R. 2014. Dehazing using color-lines. *ACM transactions on graphics (TOG)*, 34(1): 1–14.

Fu, X.; Zhuang, P.; Huang, Y.; Liao, Y.; Zhang, X.-P.; and Ding, X. 2014. A retinex-based enhancing approach for single underwater image. In *2014 IEEE international conference on image processing (ICIP)*, 4572–4576. Ieee.

Guo, C.-L.; Yan, Q.; Anwar, S.; Cong, R.; Ren, W.; and Li, C. 2022a. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5812–5820.

Guo, Y.-C.; Kang, D.; Bao, L.; He, Y.; and Zhang, S.-H. 2022b. Nerfren: Neural radiance fields with reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18409–18418.

He, K.; Sun, J.; and Tang, X. 2010. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12): 2341–2353.

Jin, Z.; Chen, S.; Feng, H.; Xu, Z.; Li, Q.; and Chen, Y. 2023. Reliable Image Dehazing by NeRF. *arXiv preprint arXiv:2303.09153*.

Ju, M.; Ding, C.; Guo, Y. J.; and Zhang, D. 2019. IDGCP: Image dehazing based on gamma correction prior. *IEEE Transactions on Image Processing*, 29: 3104–3118.

Lee, D.; Lee, M.; Shin, C.; and Lee, S. 2023. Dp-nerf: Deblurred neural radiance field with physical scene priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12386–12396.

Lee, D.-H.; et al. 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, 896. Atlanta.

Levy, D.; Peleg, A.; Pearl, N.; Rosenbaum, D.; Akkaynak, D.; Korman, S.; and Treibitz, T. 2023. Seathru-nerf: Neural radiance fields in scattering media. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 56–65.

Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; and Tao, D. 2019. An underwater image enhancement benchmark dataset and beyond. *IEEE transactions on image processing*, 29: 4376–4389.

Li, T.; Li, L.; Wang, W.; and Feng, Z. 2023. DehazingNeRF: neural radiance fields from hazy images. *arXiv preprint arXiv:2304.11448*.

Liang, Z.; Ding, X.; Wang, Y.; Yan, X.; and Fu, X. 2021. GUDCP: Generalization of underwater dark channel prior for underwater image restoration. *IEEE transactions on circuits and systems for video technology*, 32(7): 4879–4884.

Liu, J.; Liu, W.; Sun, J.; and Zeng, T. 2021. Rank-one prior: Toward real-time scene recovery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14802–14810.

Liu, Y.; Huang, T.; Dong, W.; Wu, F.; Li, X.; and Shi, G. 2023. Low-Light Image Enhancement with Multi-stage Residue Quantization and Brightness-aware Attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12140–12149.

Liu, Y.; Pan, J.; Ren, J.; and Su, Z. 2019. Learning deep priors for image dehazing. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2492–2500.

Ma, L.; Li, X.; Liao, J.; Zhang, Q.; Wang, X.; Wang, J.; and Sander, P. V. 2022. Deblur-nerf: Neural radiance fields from blurry images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12861–12870.

Max, N. 1995. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2): 99–108.

McCartney, E. J. 1976. Optics of the atmosphere: scattering by molecules and particles. *New York*.

Mildenhall, B.; Hedman, P.; Martin-Brualla, R.; Srinivasan, P. P.; and Barron, J. T. 2022. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16190–16199.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Peng, C.; and Chellappa, R. 2023. PDRF: progressively deblurring radiance field for fast scene reconstruction from blurry images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 2029–2037.

Pizer, S. M.; Amburn, E. P.; Austin, J. D.; Cromartie, R.; Geselowitz, A.; Greer, T.; ter Haar Romeny, B.; Zimmerman, J. B.; and Zuiderveld, K. 1987. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3): 355–368.

Ramazzina, A.; Bijelic, M.; Walz, S.; Sanvito, A.; Scheuble, D.; and Heide, F. 2023. ScatterNeRF: Seeing Through Fog with Physically-Based Inverse Neural Rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 17957–17968.

Sethuraman, A. V.; Ramanagopal, M. S.; and Skinner, K. A. 2023. Waternerf: Neural radiance fields for underwater scenes. In *OCEANS 2023-MTS/IEEE US Gulf Coast*, 1–7. IEEE.

Shen, H.; Zhao, Z.-Q.; Zhang, Y.; and Zhang, Z. 2023. Mutual information-driven triple interaction network for efficient image dehazing. In *Proceedings of the 31st ACM International Conference on Multimedia*, 7–16.

Song, Y.; He, Z.; Qian, H.; and Du, X. 2023. Vision transformers for single image dehazing. *IEEE Transactions on Image Processing*, 32: 1927–1941.

Tang, K.; Yang, J.; and Wang, J. 2014. Investigating haze-relevant features in a learning framework for image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2995–3000.

Teigen, A. L.; Yip, M.; Hamran, V. P.; Skui, V.; Stahl, A.; and Mester, R. 2023. Removing Adverse Volumetric Effects From Trained Neural Radiance Fields. *arXiv preprint arXiv:2311.10523*.

Tseng, P.; and Yun, S. 2009. A coordinate gradient descent method for nonsmooth separable minimization. *Mathematical Programming*, 117: 387–423.

Verbin, D.; Hedman, P.; Mildenhall, B.; Zickler, T.; Barron, J. T.; and Srinivasan, P. P. 2022. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5481–5490. IEEE.

Wang, C.; Wu, X.; Guo, Y.-C.; Zhang, S.-H.; Tai, Y.-W.; and Hu, S.-M. 2022a. Nerf-sr: High quality neural radiance fields using supersampling. In *Proceedings of the 30th ACM International Conference on Multimedia*, 6445–6454.

Wang, P.; Zhao, L.; Ma, R.; and Liu, P. 2023. Bad-nerf: Bundle adjusted deblur neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4170–4179.

Wang, Z.; Li, L.; Shen, Z.; Shen, L.; and Bo, L. 2022b. 4k-nerf: High fidelity neural radiance fields at ultra high resolutions. *arXiv preprint arXiv:2212.04701*.

Wright, S. J. 2015. Coordinate descent algorithms. *Mathematical programming*, 151(1): 3–34.

Wu, R.-Q.; Duan, Z.-P.; Guo, C.-L.; Chai, Z.; and Li, C. 2023. Ridcp: Revitalizing real image dehazing via high-quality codebook priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22282–22291.

Yu, H.; Li, X.; Lou, Q.; Lei, C.; and Liu, Z. 2020. Underwater image enhancement based on DCP and depth transmission map. *Multimedia Tools and Applications*, 79: 20373–20390.

Zhang, T.; and Johnson-Roberson, M. 2023. Beyond NeRF Underwater: Learning neural reflectance fields for true color correction of marine imagery. *IEEE Robotics and Automation Letters*.

Zhao, C.; Cai, W.; Dong, C.; and Hu, C. 2024. Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8281–8291.

Zheng, Y.; Zhan, J.; He, S.; Dong, J.; and Du, Y. 2023. Curricular contrastive regularization for physics-aware single image dehazing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5785–5794.

Zhou, J.; Liang, T.; He, Z.; Zhang, D.; Zhang, W.; Fu, X.; and Li, C. 2023. WaterHE-NeRF: Water-ray Tracing Neural Radiance Fields for Underwater Scene Reconstruction. *arXiv preprint arXiv:2312.06946*.

Zhu, C.; Wan, R.; and Shi, B. 2022. Neural transmitted radiance fields. *Advances in Neural Information Processing Systems*, 35: 38994–39006.

Zhu, Q.; Mai, J.; and Shao, L. 2015. A fast single image haze removal algorithm using color attenuation prior. *IEEE transactions on image processing*, 24(11): 3522–3533.