ScatterNeRF:
Seeing Through Fog with Physically-Based Inverse Neural Rendering

Andrea Ramazzina[1] Mario Bijelic[2] Stefanie Walz[1] Alessandro Sanvito[1] Dominik Scheuble[1] Felix Heide[2]
[1]Mercedes-Benz [2]Princeton University

Abstract

Vision in adverse weather conditions, whether it be snow, rain, or fog is challenging. In these scenarios, scattering and attenuation severly degrades image quality. Handling such inclement weather conditions, however, is essential to operate autonomous vehicles, drones and robotic applications where human performance is impeded the most. A large body of work explores removing weather-induced image degradations with dehazing methods. Most methods rely on single images as input and struggle to generalize from synthetic fully-supervised training approaches or to generate high fidelity results from unpaired real-world datasets. With data as bottleneck and most of today's training data relying on good weather conditions with inclement weather as outlier, we rely on an inverse rendering approach to reconstruct the scene content. We introduce ScatterNeRF, a neural rendering method which adequately renders foggy scenes and decomposes the fog-free background from the participating media – exploiting the multiple views from a short automotive sequence without the need for a large training data corpus. Instead, the rendering approach is optimized on the multi-view scene itself, which can be typically captured by an autonomous vehicle, robot or drone during operation. Specifically, we propose a disentangled representation for the scattering volume and the scene objects, and learn the scene reconstruction with physics-inspired losses. We validate our method by capturing multi-view In-the-Wild data and controlled captures in a large-scale fog chamber. Our code and datasets are available at https://light.princeton.edu/scatternerf.

1. Introduction

Imaging and scene understanding in the presence of scattering media, such as fog, smog, light rain and snow, is an open challenge for computer vision and photography. As rare out-of-distribution events that occur based on geography and region [8], these weather phenomena can drastically reduce the image quality of the captured intensity images, reducing local contrast, color reproduction, and image resolution [8]. A large body of existing work has investigated methods for dehazing [57, 5, 49, 29, 73, 77] with the most successful methods employing learned feed-forward models [57, 5, 49, 29, 73]. Some methods [49, 5, 35] use synthetic data and full supervision, but struggle to overcome the domain gap between simulation and real world. Acquiring paired data in real world conditions is challenging and existing methods either learn natural image priors from large unpaired datasets [74, 73], or they rely on cross-modal semi-supervision to learn to separate atmospheric effects from clear RGB intensity [57]. Unfortunately, as the semi-supervised training cues are weak compared to paired supervised data, these methods often fail to completely separate atmospheric scatter from clear image content, especially at long distances. The problem of predicting clear images in the presence of haze is an open challenge, and notably harsh weather also results in severely impaired human vision – a major driver behind fatal automotive accidents [4].

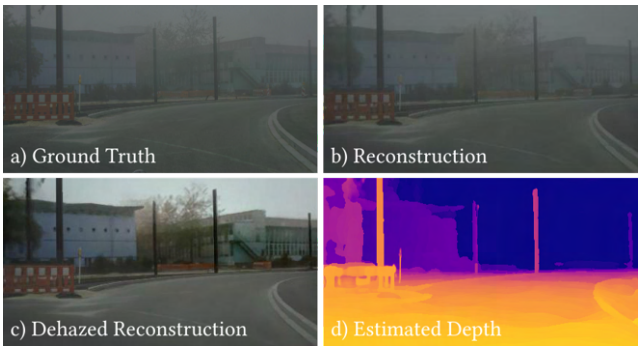

Figure 1: ScatterNeRF produces accurate renderings for scenes with volumetric scattering (b). By learning a disentangled representation of participating media and clear scene, the proposed method is able to recover dehazed scene content (c) with accurate depth (d).

As the distribution of natural scenes with participating media is long-tailed in typical training datasets [14, 60, 19, 18, 8], this also makes training and evaluation of computer vision tasks that operate on RGB streams in bad weather challenging. For supervised approaches to scene understanding tasks, these "edge" scenarios often directly re-

sult in failure cases, including detection [8], depth estimation [20], and segmentation [52]. To tackle weather-based dataset bias, existing methods have proposed augmentation approaches that either simulate atmospheric weather effects on clear images [52, 65] or they employ fully synthetic simulation to generate physically-based adverse weather scenarios [13, 23, 52]. Unfortunately, both directions cannot compete with supervised training data, either due to the domain gap between real and synthetic data, or, as a result of an approximate physical forward model [65].
As such, the capability of both physically accurate modeling and separating scattering in participating media is essential for imaging and scene understanding tasks.

In this work, we depart from both feed-forward dehazing methods and fully synthetic training data, and we address this challenge as an inverse rendering problem. Instead of predicting clean images from RGB frames, we propose to learn a neural scene representation that explains foggy images with a physically accurate forward rendering process. Once this representation is fitted to a scene, this allows us to render *novel views with real-world physics-based scattering*, *disentangle appearance and geometry without scattering* (i.e., reconstruct the dehazed scene), and *estimate physical scattering parameters* accurately. To be able to optimize a scene representation efficiently, we build on the large body of neural radiance field methods [45, 75, 69, 70, 72, 47, 32, 9, 51, 58, 10, 76, 7, 66]. While existing NeRF methods assume peaky ray termination distributions and free-space propagation, we propose a forward model that can accurately model participating media. As an inductive bias, the scene representation, by design, separates learning the clear 3D scene and the participating media. We validate that the proposed method accurately models foggy scenes in real-world and controlled scenes, and we demonstrate that the disentangled scene intensity and depth outperform existing dehazing and depth estimation methods in diverse driving scenes.

Specifically, we make the following contributions:

- We propose a method to learn a disentangled representation of the participating media by introducing the Koschmieder scattering model into the volume rendering process.
- Our approach adds a single MLP used to model the scattering media proprieties and does not require any additional sampling or other procedures, making it a lightweight framework in terms of both computation and memory consumption.
- We validate that our method learns a physics-based representation of the scene and enables control over its hazed appearance. We confirm this using real data captured in both controlled and In-the-Wild settings.

2. Related Work

Imaging and Vision in Participating Media. As real world data in participating media is challenging to capture [8, 36, 2, 1], a large body of work introduces simulation techniques for snowfall [43], rainfall [24, 23], raindrops on the windshield [68], blur [33], night [53] and fog [57, 52, 39, 17]. Using this data, existing methods investigate pre-processing approaches using dehazing [57, 5, 49, 29, 73, 77, 25, 35], deraining [27, 16] and desnowing [43]. Early works on image dehazing as [25, 64] explore image statistics and physical models to estimate the airlight introduced by the fog scattering volume and invert the Koschmieder fog model. Later, CNN approaches [35] and [29] learn to predict the airlight and transmission, with the same goal of inverting the Koschmieder model. However, this disjoint optimization can lead to error accumulation. Hence, [49, 5, 57, 73, 22] model the fog removal through a neural network learned end-to-end. As such, existing methods differ substantially in network structure and learning methodology. For example, some methods rely on GAN architectures [49, 57], transformer-based backbones [22] or encoder-decoder structures [5]. For model-learning, the approaches apply semi-/self-supervised [73, 57], fully supervised [49, 5] and test time optimization techniques [42, 29]. All of these methods have in common that they do not explore multiple views to reconstruct a clear image. To overcome this methodological limitation we introduce a novel multi-view dataset in hazy conditions in Sec. 4 and explore reconstruction from multiple views for optimal image reconstruction through neural rendering approaches. Most similar to our method are the approaches from Sethuraman et al. [56] and from Levy et al. [34] that reconstruct scenes in underwater conditions which requires tackling strong color aberrations and specular reflections specific to the underwater domain.

Neural Rendering Methods for Large Scale Scenes A rapidly growing body of work is capable of representing unbounded outdoor scenes, such as the ones tackled in this work, with rich detail both in close and far range. NeRF++ [76] achieves this by using two NeRFs, one to model the foreground and one for the background scene. DONeRF [46] warps the space by radial distortion to bring the background closer. Recently, [7] has tackled this problem by proposing a non-linear parametrization technique suitable for the mip-NeRF algorithm [6]. For very large-scale scenes, too big to be fitted by a single NeRF, several works [62] [66] have explored the idea of learning multiple NeRFs for subparts of the scene.

Clear Scene Priors Given the under-constrained nature of scene reconstruction from a sparse set of views, novel views rendered by NeRFs are frequently afflicted by floating artifacts [7, 12] or not able to properly generalize to novel

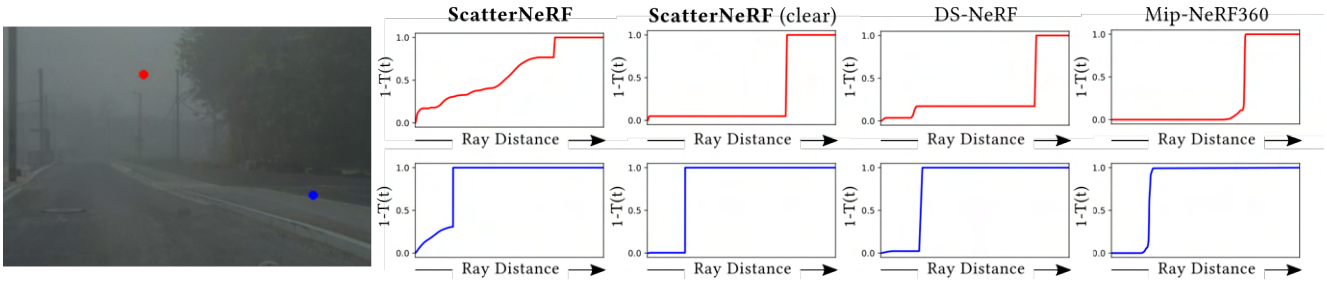

Figure 2: We show the ray termination distribution along two cast rays for our approach and two references NeRF models. The regularization methods proposed in DS-NeRF [12] and mip-NeRF-360 [7] represent the accumulated transmittance T as a step function, whereas ScatterNeRF models the scattering process.

views and hence fail to render images from unseen poses correctly [30, 28]. To tackle this issue, several works have recently proposed the introduction of different regularization techniques. In [30], an entropy-based loss is introduced in order to enforce a sparsity constraint over the scene. Analogously, recent models exploiting an estimated depth as training cue [12, 50] implicitly enforce a sparsity constrain of the ray termination distributions by penalizing non-zero probabilities lying far from the prior estimated depth. Mip-NeRF360 [7] relies on a regularization technique aimed at encouraging a unimodal peaky distribution for the termination probability of a ray. While such methods work well for clear scenes, they are based on the assumption that most of the scene density is null, except for where there are solid objects. As such prior is not applicable in the presence of a participating media in the air, such approaches are not suitable for scenes with scattering media.

We propose to learn a separate representation of the clear scene and the participating media. This allows us to potentially use any of the above-mentioned regularization technique on the clear scene model without compromising on the hazed scene reconstruction.

3. Disentangled Scattering Neural Radiance Fields

ScatterNeRF has five integral parts, namely the underlying physical model in Sec. 3.1, the formulation of the neural radiance field in Sec. 3.2, the formulation of the loss functions in Sec. 3.3, the details on ray sampling in Sec. 3.4 and implementation in Sec. 3.5. We describe these components in the following.

3.1. Physical Scattering Model

Large scattering volumes can be approximated by the Koschmieder model [31]. For each pixel, we model the attenuation representing the lost intensity due to scattered direct light of an object and the contribution of the airlight $\mathbf{c}_p$ caused by ambient light scattered towards the observer,

$$\mathbf{C}_F = l\mathbf{C}_c + (1-l)\mathbf{C}_p, \tag{1}$$

with l being the transmission, $\mathbf{C}_F$ corresponding to the observed pixel value and $\mathbf{C}_c$ the clear scene. The transmission l can be computed from the attenuation coefficient σ_p and the depth D,

$$l = \exp(-\sigma_p D). \tag{2}$$

In existing methods [52] both parameters σ_p and airlight $\mathbf{c}_p$ are globally constant. Koschmieder's model is equivalent to a volume rendering NeRF [45], Eq. (3), with scattering density σ_p and airlight $\mathbf{C}_p$ set constant, and ray integration from t_n to a maximum distance t_f. For simplicity of notation, we omit the viewing direction $\mathbf{d}$. Starting with the forward model

$$\mathbf{C}_F(\mathbf{r}) = \int_{t_n}^{t_f} T_F(t)\sigma_F(\mathbf{r}(t))\mathbf{c}_F(\mathbf{r}(t))dt, \tag{3}$$

and assuming two disjoint additive volume densities σ_c for the scene and σ_p for the scattering media ($\sigma_F = \sigma_p + \sigma_c$), the volume rendering equation can be formulated as

$$\mathbf{C}_F(\mathbf{r}) = l(\mathbf{r})\mathbf{C}_c(\mathbf{r}) + (1 - l(\mathbf{r}))\,\mathbf{C}_p, \tag{4}$$

where the integrated object color $\mathbf{C}_c$ is defined with emitted color $\mathbf{c}_c(\mathbf{r})$ at position $\mathbf{r}$,

$$\mathbf{C}_c(\mathbf{r}) = \int_{D}^{t_f} T_c(t)\sigma_c(\mathbf{r}(t))c_c(\mathbf{r}(t))dt. \tag{5}$$

The transmission then is dependent on $\mathbf{r}$ leading to $\exp(-\sigma_p D)$. Here, the subscript denotes if the clear scene object c or participating media p is modeled. Finally, the value of the foggy pixel $\mathbf{C}_F(\mathbf{r})$ can be estimated by integrating $\mathbf{c}_F(\mathbf{r}(t))$ along one camera ray vector $\mathbf{r}$. We denote $\mathbf{r}_i = [\mathbf{x}_i; \mathbf{d}_i] \in \mathbb{R}^5$ consisting of the 3D position $\mathbf{x}$ and direction $\mathbf{d}$.

Next, we can relax the constraints on σ_p and $\mathbf{c}_p$, and allow them to approximate arbitrary values. This results in,

$$\begin{aligned}\mathbf{C}_F(\mathbf{r}) = \int_{t_n}^{t_f} T_F(t)(&\sigma_p(\mathbf{r}(t))\mathbf{c}_p(\mathbf{r}(t)) \\ &+ \sigma_c(\mathbf{r}(t))\mathbf{c}_c(\mathbf{r}(t)))dt,\end{aligned} \tag{6}$$

with,

$$T_F(t) = \exp\left(-\int_{t_n}^{t}(\sigma_c(\mathbf{r}(s)) + \sigma_p(\mathbf{r}(s)))ds\right), \tag{7}$$

$$T_F(t) = T_p(t)T_c(t), \tag{8}$$

which can be expressed as

$$\mathbf{C}_F(\mathbf{r}) = \int_{t_n}^{t_f} T_c(t) \underbrace{\left(T_p(t)\sigma_p(\mathbf{r}(t))\mathbf{c}_p(\mathbf{r}(t))\right)}_{\text{Fog Contribution}} + T_p(t) \underbrace{\left(T_c(t)\sigma_c(\mathbf{r}(t))\mathbf{c}_c(\mathbf{r}(t))\right)}_{\text{Clear scene contribution}} dt. \quad (9)$$

3.2. Neural Radiance Model

Further simplifications of Eq. (3) can be found by solving the integral through numerical quadrature needed for the discrete neural forward network. The numerical quadrature leads to,

$$\mathbf{C}_F(\mathbf{r}) = \sum_i^N w_F(\mathbf{r}(t_i))\mathbf{c}_F(\mathbf{r}(t_i)). \quad (10)$$

$$w_F(\mathbf{r}_i) = T_F(\mathbf{r}_i)(1 - \exp\left((\sigma_p(\mathbf{r}_i) + \sigma_c(\mathbf{r}_i))\,\delta_j\right), \quad (11)$$

$$T_F(\mathbf{r}_i) = \exp\left(-\sum_{j=1}^{i-1}(\sigma_p(\mathbf{r}_i) + \sigma_c(\mathbf{r}_i))\,\delta_i\right), \quad (12)$$

$$\delta_i = t_{i+1} - t_i, \text{ and} \quad (13)$$

$$\mathbf{c}_F(\mathbf{r}_i, \mathbf{d}) = \frac{\sigma_c(\mathbf{r}_i)\mathbf{c}_c(\mathbf{r}_i) + \sigma_p(\mathbf{r}_i)\mathbf{c}_p(\mathbf{r}_i)}{\sigma_p(\mathbf{r}_i) + \sigma_c(\mathbf{r}_i)}. \quad (14)$$

To facilitate the learning of two independent volume representations, we model each part independently by one NeRF. The parameters $\sigma_i, \mathbf{c}_i$ for $i \in \{c, p\}$ are predicted by a multi-layer perceptron (MLP) as,

$$\mathbf{c}_i, \sigma_i = f_i\left(\gamma(\mathbf{x})\right). \quad (15)$$

For the clear scene NeRF we adopt a similar strategy as in [45] and optimize simultaneously two MLPs, $f_{c_{coarse}}$ and $f_{c_{fine}}$ using the same loss formulation but different sampling procedure, as detailed in Sec. 3.3 and Sec. 3.4. The coordinates are encoded by the function γ following [45, 63]. This learning disentanglement of the scene and scatter representation allows us to render scenes with different fog densities by scaling σ_p or even dehaze the image entirely by setting $\sigma_p = 0$.

For the dehazing task the image can be rerendered with $\sigma_p = 0$ and the formulation reduces to the scene model only leading to,

$$\mathbf{C}_F(\mathbf{r}) = \mathbf{C}_c(\mathbf{r}) = \sum_i^N w_c(\mathbf{r}(t_i))\mathbf{c}_c(\mathbf{r}(t_i)), \quad (16)$$

$$w_c(\mathbf{r}(t_i)) = T_c(t_i)(1 - \exp\left(\sigma_c(\mathbf{r}(t_i))\delta_i\right), \quad (17)$$

$$T_c(t_i) = \exp\left(-\sum_{j=1}^{i-1}(\sigma_c(\mathbf{r}(t_j))\delta_j\right). \quad (18)$$

3.3. Training Supervision

To learn the neural forward model we supervise the reconstructed images with a pixel loss between predicted and training frames $\mathcal{L}_{rgb}$, align the observable airlight $\mathcal{L}_A$, supervise the scene depth with $\mathcal{L}_D$, enforce discrete clear scene volumetric density $\mathcal{L}_{ec}$ and enforce the scattering density to be disjoint from the scene objects in $\mathcal{L}_{eF}$. In the following, network predictions are marked with a hat and ground-truth values are marked with a bar.

Color Supervision The image loss between ground-truth $\bar{\mathbf{C}}_F$ and reconstructed haze images $\hat{\mathbf{C}}_F$ is used for direct supervision as

$$L_{rgbF} = \mathbb{E}_{\mathbf{r}}\left[||\hat{\mathbf{C}}_F(\mathbf{r}) - \bar{\mathbf{C}}_F(\mathbf{r})||_2^2\right]. \quad (19)$$

Airlight Color Supervision As discussed in Sec. 3.1, the model separation allows us to supervise the airlight directly. We minimize the variance of the predicted $\hat{\mathbf{c}}_p$ with respect to a ground-truth airlight $\bar{\mathbf{c}}_p$ estimated following [64]. The target airlight $\bar{\mathbf{c}}_p$ is computed as follows,

$$\bar{\mathbf{c}}_p(\mathbf{r}) = \frac{z^2(\mathbf{r})I_F(\mathbf{r}) + \lambda\mathbf{c}_p^0(\mathbf{r})}{z^2(\mathbf{r}) + \lambda}. \quad (20)$$

Here I_F is the relative luminance of the hazed image estimated as $I_F(\mathbf{r}) = \xi \cdot \text{lin}(\bar{\mathbf{C}}_F(\mathbf{r}))$, that is a linear combination of the linearized RGB values [3] obtained by decompanding the color image by applying $\text{lin}(\bar{\mathbf{C}}_F)$ [61]. $\mathbf{c}_p^0$ is an initial global constant airlight estimate computed with the dark channel prior following [25], $z = 1/(l-1)$ and λ is a weighting factor. The total loss can be written as,

$$L_A = \mathbb{E}_{\mathbf{r}}\left[||\hat{\mathbf{c}}_p(\mathbf{r}) - \bar{\mathbf{c}}_p(\mathbf{r})||_2^2\right]. \quad (21)$$

Clear Scene Entropy Minimization To regularize f_c, we follow [12] according to which the rays cast in the scene have peaky unimodal ray termination distributions. Thus, we add a loss to minimize its distribution entropy,

$$L_{ec} = \mathbb{E}_r\left[-\sum_i \hat{w}_{c_i}(\mathbf{r}) \cdot \log\left(\hat{w}_{c_i}(\mathbf{r})\right)\right]. \quad (22)$$

Foggy Scene Entropy Maximization Analogously to Eq. (22), we regularize f_p. Thereby, $\hat{\sigma}_p$ is fitted in semi-supervised fashion during the optimization. This allows us to model fog inhomogeneities, for example close to hot surfaces. To achieve this goal we apply an entropy-based loss which allows the network f_p to learn a spatially-varying media density. Based on the assumption of almost-uniformity

Figure 3: Examples of our proposed In-the-Wild and controlled environment dataset. The figure demonstrates the diversity of the scenes and the fog densities.

for extended fog volume, we enforce that this has to be represented by the fog volume density $\hat{\sigma}_F$ through maximizing the entropy as follows,

$$L_{eF} = \mathbb{E}_r \left[\sum_i \tilde{\alpha}_{F_i}(\mathbf{r}) \cdot \log(\tilde{\alpha}_{F_i}(\mathbf{r})) \right], \tag{23}$$

where $\tilde{\alpha}_{F_i} = \frac{\hat{\alpha}_{F_i}}{\sum_j \hat{\alpha}_{F_j}}$ and $\hat{\alpha}_{F_i} = 1 - exp((\hat{\sigma}_{p_i} + \hat{\sigma}_{c_i})\delta_i)$. The entropy maximization relies on the scene volume density $\hat{\sigma}_c$ to disentangle both distributions and not only distribute the fog volume density $\hat{\sigma}_F$ throughout the scene. We minimize this loss only for $\hat{\sigma_p}$, not $\hat{\sigma_c}$.

Estimated Depth Supervision To supervise the scene representation further we supervise the scene depth $\hat{D}_c$ similar to [50], through the depth $\bar{D}$ estimated from the stereo sensor setup. Thereby, the depth can be estimated as follows,

$$\hat{D}_c(\mathbf{r}) = \sum_i^N \hat{w}_c(\mathbf{r}(t_i))t_i \tag{24}$$

which leads to,

$$L_{depth} = \mathbb{E}_r \left[||\hat{D}_c(\mathbf{r}) - \bar{D}(\mathbf{r})||_2^2 \right]. \tag{25}$$

We use a pretrained RGB stereo algorithm [37] to compute $\hat{D}$.

Total Training Loss Combining the five losses, we obtain the following loss formulation,

$$L_{tot} = \psi_1 L_{rgbF} + \psi_2 L_A + \psi_3 L_{ec} + \psi_4 L_{eF} + \psi_5 L_{depth}. \tag{26}$$

Where $\psi_{1,\dots,5}$ are the loss weights, provided in the Supplementary Material.

3.4. Sampling

We follow the hierarchical volume sampling strategy proposed in [45] to sample the query points for the radiance field networks. However, instead of sampling across the whole volume density we adapt the approach to our decomposed volume densities σ_c and σ_F. As σ_F is regularized to be approximately constant across the scene, the resampling procedure is not going to be performed using the scene weights $w_F = T_F(1 - \exp(\sigma_F \delta))$ but rather using the clear scene $w_c = T_c(1 - \exp(\sigma_c \delta))$. We apply this approach to follow an importance sampling and reconstruct scene objects by sampling close to object boundaries.

3.5. Implementation Details

We train for 250'000 steps and a batch size of 4096 rays. As optimizer we use ADAMW [44] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, learning rate $5 \cdot 10^{-4}$ and weight decay factor of 10^{-2}. Our code implementation is based on Pytorch [48] and we train on four NVIDIA RTX A6000. The NeRF MLPs $f_{c_{coarse}}$ and $f_{c_{fine}}$ follow Mildenhall et al. [45], while we use fewer hidden layers for f_p. The network architecture and other hyper-parameters are listed in the Supplementary Material.

4. Dataset

To evaluate the proposed method, we collect both an automotive In-the-Wild foggy dataset and a controlled fog dataset. Example captures illustrating the dataset are shown in Fig. 3. In total, we collect 2678 In-the-Wild foggy images throughout nine different scenarios. The sensor set for the In-the-Wild dataset consists of an Aptina-0230 stereo camera pair and a Velodyne HDL64S3D laser scanner. Camera poses are estimated with the hierarchical localization approach [54, 55], a structure from motion pipeline optimized for robustness to changing conditions. The training and testing split was done by randomly choosing 30% of the images as test set.Each scene contains between 150 and 300 images. Ground-truth depth data is estimated with the stereo-camera method described in [38]. The controlled fog dataset is captured in a fog chamber where fog with varying visibilities can be generated. We capture 903 images in a large-scale fog chamber with clear ground-truth and two different fog densities. Further ground-truth depths are captured through a Leica ScanStation P30 laser scanner (360°/290° FOV, 1550 nm, with up to 1M points per second, up to 8" angular accuracy, and 1.2mm + 10 parts per million (ppm) range accuracy). Each point cloud consists of approximately 157M points and we accumulate multiple point clouds from different positions to reduce occlusions and increase resolution.

5. Assessment

Next, we validate the proposed method by ablating the different components, confirming their effectiveness, assessing the quality of scene reconstructions in foggy scenes, the decomposition into dehazed scene content, and the scene depth reconstruction.

Method	Tram Station			Farm			Intersection			Suburb			Speed Control		
	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑
Aug-NeRF [11]	0.26	39.18	0.964	0.316	39.83	0.968	0.275	40.38	0.971	0.259	40.18	0.977	0.277	40.54	0.974
DS-NeRF [12]	0.265	39.99	0.967	0.332	41.27	0.97	0.273	42.19	0.977	0.271	42.89	0.978	0.281	42.8	0.976
DVGO [59]	0.404	35.46	0.923	0.441	37.49	0.93	0.4	39.35	0.955	0.36	39.68	0.963	0.382	39.16	0.958
Mip-NeRF [6]	0.333	40.29	0.964	0.337	40.52	0.968	0.271	41.72	0.975	0.27	41.23	0.977	0.28	41.72	0.975
Mip-NeRF360 [7]	0.222	40.92	0.971	**0.28**	41.25	**0.973**	**0.228**	42.17	0.976	0.244	41.91	0.977	**0.247**	42.2	0.977
NeRF [45]	0.278	39.06	0.964	0.354	40.86	0.968	0.293	41.84	0.974	0.271	41.6	0.978	0.293	42.39	0.976
NeRF++ [76]	0.288	39.54	0.962	0.36	41.07	0.966	0.306	42.11	0.972	0.293	42.07	0.976	0.304	42.41	0.974
Plenoxels [15]	0.476	30.9	0.91	0.498	32.28	0.939	0.478	29.93	0.934	0.486	28.29	0.928	0.489	29.34	0.934
Ref-NeRF [67]	0.4	36.64	0.94	0.377	39.85	0.966	0.377	40.53	0.964	0.343	40.47	0.972	0.366	41.13	0.967
ScatterNeRF	**0.22**	**41.45**	**0.975**	0.299	**42.57**	0.972	0.235	**44.58**	**0.98**	**0.234**	**44.44**	**0.981**	0.253	**43.88**	**0.978**

Method	Road Fork			Adds			Construction			Countryside			Average In-the-Wild		
	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑
Aug-NeRF [11]	0.248	41.29	0.979	0.242	41.75	0.981	0.302	38.99	0.964	0.286	40.74	0.978	0.274	40.32	0.973
DS-NeRF [12]	0.273	42.59	0.977	0.252	43.22	0.982	0.318	40.94	0.964	0.29	43.6	0.98	0.284	42.16	0.975
DVGO [59]	0.376	38.51	0.947	0.349	40.14	0.966	0.376	38.92	0.95	0.368	40.55	0.968	0.384	38.81	0.951
Mip-NeRF [6]	0.272	41.88	0.976	0.25	42.83	0.981	0.321	40.2	0.963	0.291	43.08	0.98	0.292	41.5	0.973
Mip-NeRF360 [7]	0.235	42.37	0.978	**0.225**	42.55	0.981	0.268	41.53	0.969	**0.241**	42.97	0.981	**0.243**	41.99	0.976
NeRF [45]	0.287	41.84	0.975	0.256	43.67	0.982	0.332	40.36	0.962	0.305	43.39	0.979	0.297	41.67	0.973
NeRF++ [76]	0.292	42.38	0.975	0.267	43.82	0.98	0.336	40.69	0.962	0.312	43.56	0.978	0.306	41.96	0.972
Plenoxels [15]	0.481	29.85	0.933	0.456	29.72	0.943	0.497	29.49	0.922	0.473	29.47	0.938	0.482	29.92	0.931
Ref-NeRF [67]	0.382	39.61	0.962	0.376	40.81	0.967	0.365	40.04	0.96	0.339	42.81	0.976	0.369	40.21	0.964
ScatterNeRF	**0.231**	**44.45**	**0.981**	0.228	**45.27**	**0.983**	**0.255**	**42.96**	**0.973**	0.265	**44.64**	**0.982**	0.247	**43.8**	**0.978**

Table 1: Quantitative comparison of the proposed ScatterNeRF and state-of-the-art methods on In-the-Wild sequences. Best results in each category are in **bold** and second best are underlined. Last column in the second row presents the average over all sequences.

Method	Toyota			Car Accident		
	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑
Aug-NeRF [11]	0.521	29.81	0.934	**0.48**	25.08	0.822
DS-NeRF [12]	0.535	29.52	0.926	0.502	26.71	0.853
DVGO [59]	0.613	22.65	0.801	0.607	22.22	0.847
Mip-NeRF [6]	0.513	30.42	0.937	0.611	24.88	0.87
Mip-NeRF360 [7]	0.53	30.59	**0.938**	0.623	26.59	0.878
NeRF [45]	0.505	29.81	0.937	0.614	24.97	0.877
NeRF++ [76]	0.507	30.0	0.937	0.611	25.47	**0.879**
Plenoxels [15]	0.582	20.48	0.872	0.607	22.07	0.852
Ref-NeRF [67]	**0.498**	29.11	0.933	0.616	24.7	0.864
ScatterNeRF	0.505	**30.82**	**0.938**	0.509	**27.42**	0.878

Table 2: Quantitative comparison of the proposed ScatterNeRF and state-of-the-art methods on controlled scenes. Best results in each category are in **bold** and second best are underlined.

	LPIPS	PSNR	SSIM
NeRF	0.278	39.06	0.964
NeRF + Koschmieder	0.281	39.54	0.960
NeRF + Depth [50]	0.290	40.03	0.972
ScatterNeRF w/o cleared sampling	0.219	40.70	0.975
ScatterNeRF	0.22	41.45	0.975

Table 3: Ablation study of the ScatterNeRF contribution for a subset of the In-the-Wild dataset.

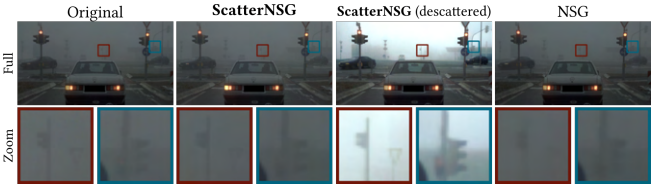

Figure 4: Qualitative comparison between Neural Scene Graphs (NSG) [47] and its combination with our proposed network (ScatterNSG).

5.1. Ablation

In order to assess the role and contribution of the different components of our framework, we conduct an ablation study whose results are presented in Tab. 3. We consider as starting point the *"NeRF"* [45]. Its PSNR on the In-the-Wild subset dataset is 39.06 dB. By adding the depth supervision Eq. (25) and Eq. (22) the model's PSNR improves by 0.97 dB. However, as shown in Fig. 2, a NeRF trained in such a way does not produce a physically accurate representation of the scene, as it does not model the participating media present in the scene, but rather represents it as a clear scene. Adding the scattering media f_p in *"ScatterNeRF w/o cleared sampling"*, together with its two losses L_A and L_{eF} leads to a PSNR of 40.7 dB. Finally, adding the sampling strategy in the full *"ScatterNeRF"* helps the model achieve better results, with a PSNR of 41.45 dB, summing up to a

SEQUENCE		ScatterNerf	PFF	D4	EPDN	ZeroScatter	ZeroRestore	MAP-net
Tram Station	FiD ↓	348.72	376.39	342.10	345.66	355.27	340.59	**301.30**
Farm		**319.95**	404.89	360.53	347.85	413.08	359.53	349.12
Intersection		**349.57**	387.09	358.63	353.27	397.71	366.82	364.04
Suburb		**283.43**	405.37	299.57	301.63	347.38	350.11	286.59
Speed Control		**270.86**	409.94	332.78	328.08	358.97	339.19	289.20
Road Fork		**284.69**	419.08	318.84	314.66	355.30	325.45	307.25
Adds		**322.19**	441.33	328.59	330.71	380.70	342.91	335.87
Construction		329.22	381.51	318.71	315.16	339.66	310.11	**308.18**
Countryside		**289.56**	435.71	326.05	321.35	337.88	321.29	303.70
Toyota	PSNR ↑	**13.47**	12.13	11.74	11.62	13.39	13.41	N/A
Car Accident		11.66	10.76	10.32	10.04	9.28	**12.02**	N/A

Table 4: Quantitative dehazing comparison on In-the-Wild dataset with FiD score and on controlled dataset with PSNR. The best results in each category are in **bold** and the second best are underlined.

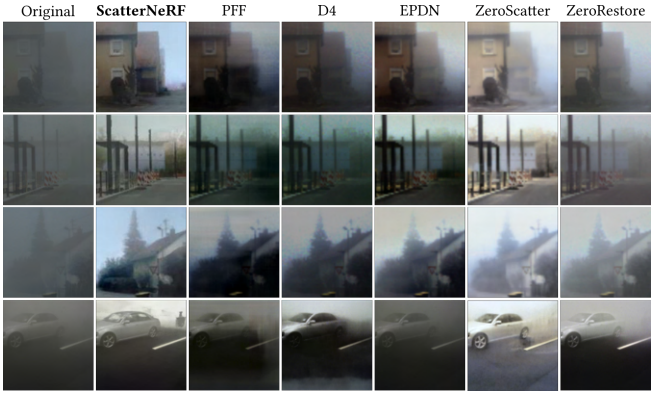

Figure 5: Qualitative comparison of dehazed images on real-world automotive measurements. The proposed ScatterNeRF enables enhanced contrast and visibility compared to state-of-the-art descattering methods.

6.12% PSNR increase over the baseline.
We also analyze an additional model, here called *"NeRF+Koschmieder"*, in which a Koschmieder model is added to the NeRF output, and we can note in Tab. 3 its limited performances due to its over-simplifications.

5.2. Foggy Scene Reconstruction

We compare our proposed method with NeRF [45], Mip-NeRF [6], DVGO [59], Plenoxel [15], Ref-NeRF [67], two methods for unbounded scenes [76, 7] and two NeRFs with auxiliary regularization [50, 11]. Qualitative results are presented in Fig. 6. The baseline methods struggle with object edges and fine structures, which our method is able to reconstruct for novel views. Our method is the only approach able to reconstruct the edges cleanly achieving to the highest PSNR scores for In-the-Wild captures and fog chamber captures. On average, the proposed method improves on average by 2.13 dB the PSNR of NeRF and by 1.64 dB to the next best model. For single sequences with high fog densities, improvements in the reconstruction of up to 2.39 db are measurable. For the SSIM metric it outperforms all other approaches except on the light foggy scene dubbed "car accident" where it seconds [15].

5.3. Generalizability

As our proposed method does not require major changes to the sampling procedure, rendering or scene representation architecture, it can be easily integrated with existing neural rendering methods. To demonstrate this, we extend NSG [47] to foggy scenes and integrate the decoupling of scene and scattering media. A qualitative comparison between the base model and the one enhanced with our framework is shown in Fig. 4, where it is possible to observe both the finer details at long distances in addition to the possibility to remove the scattering media and better reconstruct the vehicle leaving the scene. Quantitatively this improvement is reflected in an improvement of scene reconstruction PSNR, increasing from 29.78 dB to 30.67 dB. This also validates that it is possible to employ our framework *also in the presence of moving objects*, which is a common occurrence in automotive scenes. We discuss in the supplementary material other examples of the integration of ScatterNeRF in state-of-the-art neural radiance fields to augment their scope.

5.4. Dehazing

We quantify the ability of ScatterNeRF to learn a disentangle representation between objects and scattering media, and hence to render the corresponding clear scene which is effectively dehazing the foggy image. To this end, we compare against the state-of-the art dehazing methods PFF [5], D4 [73], EPND [49], ZeroScatter [57], ZeroRestore [29] and MAP-Net [71]. For the In-the-Wild scenes, we rely on the FiD score [26] to evaluate the quality of the dehazing. As ground-truth data is unavailable, this score allows us to compare the dehazed sequences with a similar clear-weather scene collected with the automotive setup described in Sec. 4. For the controlled environment dataset the PSNR can directly be evaluated by warping a nearby clear-weather image to the respective foggy image. In contrast to other dehazing methods operating on single frames, ScatterNeRF learns a consistent representation of the entire sequence. As a result, the dehazing is consistent across consecutive frames, whereas the baseline methods often are affected by flickering effects. This is also reflected in the quantitative results in Tab. 4. Here, ScatterNeRF outperforms the baseline methods on almost all sequences indicating a consistent dehazing across the entire sequence. Furthermore, the qualitative results in Fig. 5 from the controlled fog chamber dataset reveal another strength of ScatterNeRF: ScatterNeRF is the only method able to reconstruct the cart behind the vehicle as it can leverage information from the learned representation of the entire sequence. This results in a higher PSNR indicating an improved dehazing on the respective sequence. Additionally, as evident in Fig. 5, ScatterNeRF achieves visually enhanced contrast compared to existing dehazing methods.

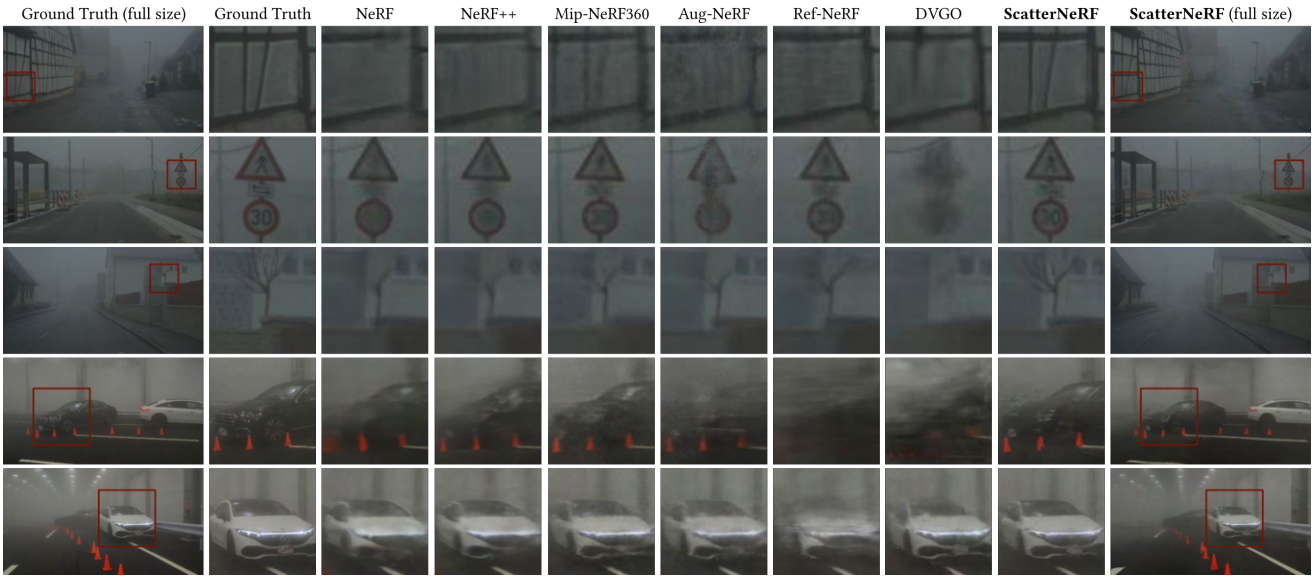

Figure 6: Qualitative comparisons of the reconstruction of foggy scenes with ScatterNeRF and state-of-the-art neural rendering methods. ScatterNeRF is able to represent the participating media much better than existing rendering methods.

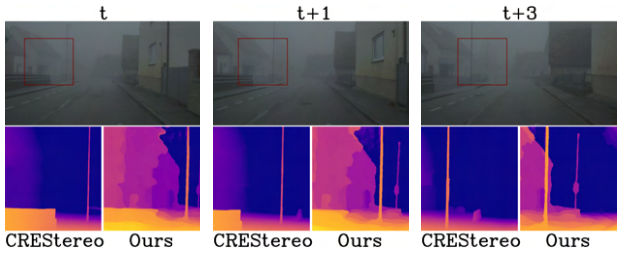

Figure 7: Qualitative results of the reconstructed depth for [37] and our method. While classical feed-forward depth estimation algorithm can not reconstruct finer details in the far back due to the fog disturbance, our method is able to render a detailed depth map thanks to its consistency constrains across all the training frames.

5.5. Depth Estimation

Depth reconstruction in foggy conditions is unreliable and performs poorly for far objects due to the loss of scene information caused by the scattering medium [21]. This can be seen in Fig. 7 where the second pole and the building in the far background is not visible in the depth estimated using [37], due to the lack of contrast in stereo matching. On the other hand, our approach enforces consistency across multiple frames and therefore achieves accurate depth also for areas distant from the cameras. Quantitatively, we compare our results with state-of-the-art RGB depth-estimation methods, namely the monocular algorithms DepthFormer [40] and BinsFormer [41], as well the stereo algorithm CREStereo [37] with a 22cm baseline. Using the LiDAR pointcloud as groundtruth, the Mean Absolute Depth Prediction Error (MAE) for our method is 3.72m, while we get 5.84m for DepthFormer, 5.85m for BinsFormer and 4.73m for CREStereo. Our algorithm outperforms the baselines by more than 21% MAE, demostrating the superiority of our approach for accurate depth reconstruction of foggy scenes.

6. Conclusion

We introduce ScatterNeRF, a neural rendering method that represents foggy scenes with disentangled representations of the participating media and scene. We model light transport in the presence of scattering with disentangled volume rendering, separately modeling the clear scene and fog, and introduce a set of physics-based losses designed to enforce the division between media and scene. Extensive experiments with both In-the-Wild and controlled scenario measurements validate the proposed approach. We demonstrate that ScatterNeRF is capable of rendering the learned scene without scattering media and can be hence used to alter or remove the haze from a sequence video, reaching quality comparable to state-of-the-art image dehazing algorithms – solely by fitting image observations without any forward neural network for dehazing or denoising.

Acknowledgments This work was supported by the AI-SEE project with funding from the FFG, BMBF, and NRC-IRA. We thank the Federal Ministry for Economic Affairs and Energy for support via the PEGASUS-family project "VVM-Verification and Validation Methods for Automated Vehicles Level 4 and 5". Felix Heide was supported by an NSF CAREER Award (2047359), a Packard Foundation Fellowship, a Sloan Research Fellowship, a Sony Young Faculty Award, a Project X Innovation Award, and an Amazon Science Research Award.

References

[1] C. Ancuti, C. O. Ancuti, R. Timofte, L. Van Gool, L. Zhang, M. Yang, V. M. Patel, H. Zhang, V. A. Sindagi, R. Zhao, X. Ma, Y. Qin, L. Jia, K. Friedel, S. Ki, H. Sim, J. Choi, S. Kim, S. Seo, S. Kim, M. Kim, R. Mondal, S. Santra, B. Chanda, J. Liu, K. Mei, J. Li, Luyao, F. Fang, A. Jiang, X. Qu, T. Liu, P. Wang, B. Sun, J. Deng, Y. Zhao, M. Hong, J. Huang, Y. Chen, E. Chen, X. Yu, T. Wu, A. Genc, D. Engin, H. K. Ekenel, W. Liu, T. Tong, G. Li, Q. Gao, Z. Li, D. Tang, Y. Chen, Z. Huo, A. Alvarez-Gila, A. Galdran, A. Bria, J. Vazquez-Corral, M. Bertalmo, H. S. Demir, O. F. Adil, H. X. Phung, X. Jin, J. Chen, C. Shan, and Z. Chen. Ntire 2018 challenge on image dehazing: Methods and results. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1004–100410, 2018. 2

[2] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 754–762, 2018. 2

[3] Matthew Anderson, Ricardo Motta, Srinivasan Chandrasekar, and Michael Stokes. Proposal for a standard default color space for the internet-srgb. In *Color Imaging Conference*, volume 6, 1996. 4

[4] Walker S. Ashley, Stephen Strader, Douglas C. Dziubla, and Alex Haberlie. Driving blind: Weather-related vision hazards and fatal motor vehicle crashes. *Bulletin of the American Meteorological Society*, 96(5):755 – 778, 2015. 1

[5] Haoran Bai, Jinshan Pan, Xinguang Xiang, and Jinhui Tang. Self-guided image dehazing using progressive feature fusion. *IEEE Transactions on Image Processing*, 31:1217 – 1229, 2022. 1, 2, 7

[6] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5855–5864, 2021. 2, 6, 7

[7] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 2, 3, 6, 7

[8] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1, 2

[9] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T Barron, Ce Liu, and Hendrik Lensch. Nerd: Neural reflectance decomposition from image collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12684–12694, 2021. 2

[10] Mark Boss, Andreas Engelhardt, Abhishek Kar, Yuanzhen Li, Deqing Sun, Jonathan T Barron, Hendrik Lensch, and Varun Jampani. Samurai: Shape and material from unconstrained real-world arbitrary image collections. *arXiv preprint arXiv:2205.15768*, 2022. 2

[11] Tianlong Chen, Peihao Wang, Zhiwen Fan, and Zhangyang Wang. Aug-nerf: Training stronger neural radiance fields with triple-level physically-grounded augmentations. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 6, 7

[12] Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. Depth-supervised nerf: Fewer views and faster training for free. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12882–12891, 2022. 2, 3, 4, 6

[13] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017. 2

[14] Scott Ettinger, Shuyang Cheng, Benjamin Caine, Chenxi Liu, Hang Zhao, Sabeek Pradhan, Yuning Chai, Ben Sapp, Charles R. Qi, Yin Zhou, Zoey Yang, Aur'elien Chouard, Pei Sun, Jiquan Ngiam, Vijay Vasudevan, Alexander McCauley, Jonathon Shlens, and Dragomir Anguelov. Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9710–9719, October 2021. 1

[15] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5501–5510, 2022. 6, 7

[16] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6):2944–2956, 2017. 2

[17] Adrian Galdran. Image dehazing by artificial multiple-exposure image fusion. *Signal Processing*, 149:135–147, 2018. 2

[18] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013. 1

[19] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 1

[20] Tobias Gruber, Mario Bijelic, Felix Heide, Werner Ritter, and Klaus Dietmayer. Pixel-accurate depth evaluation in realistic driving scenarios. In *International Conference on 3D Vision (3DV)*, 2019. 2

[21] Tobias Gruber, Mario Bijelic, Felix Heide, Werner Ritter, and Klaus Dietmayer. Pixel-accurate depth evaluation in realistic driving scenarios. In *2019 International Conference on 3D Vision (3DV)*, pages 95–105, 2019. 8

[22] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5812–5820, June 2022. 2

[23] Shirsendu Sukanta Halder, Jean-François Lalonde, and Raoul de Charette. Physics-based rendering for improving robustness to rain. In *ICCV*, 2019. 2

[24] S. Hasirlioglu and A. Riener. A general approach for simulating rain effects on sensor data in real and virtual environments. *IEEE Transactions on Intelligent Vehicles*, pages 1–1, 2019. 2

[25] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010. 2, 4

[26] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 6626–6637, 2017. 7

[27] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. Depth-attentional features for single-image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2

[28] Ajay Jain, Matthew Tancik, and Pieter Abbeel. Putting nerf on a diet: Semantically consistent few-shot view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5885–5894, 2021. 3

[29] Aupendu Kar, Sobhan Kanti Dhara, Debashis Sen, and Prabir Kumar Biswas. Zero-shot single image restoration through controlled perturbation of koschmieder's model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16205–16215, June 2021. 1, 2, 7

[30] Mijeong Kim, Seonguk Seo, and Bohyung Han. Infonerf: Ray entropy minimization for few-shot neural volume rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12912–12921, 2022. 3

[31] Harald Koschmieder. Theorie der horizontalen sichtweite. *Beitrage zur Physik der freien Atmosphare*, pages 33–53, 1924. 3

[32] Abhijit Kundu, Kyle Genova, Xiaoqi Yin, Alireza Fathi, Caroline Pantofaru, Leonidas J Guibas, Andrea Tagliasacchi, Frank Dellaert, and Thomas Funkhouser. Panoptic neural fields: A semantic object-aware neural scene representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12871–12881, 2022. 2

[33] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018. 2

[34] Deborah Levy, Amit Peleg, Naama Pearl, Dan Rosenbaum, Derya Akkaynak, Simon Korman, and Tali Treibitz. Seathru-nerf: Neural radiance fields in scattering media, 2023. 2

[35] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE international conference on computer vision*, pages 4770–4778, 2017. 1, 2

[36] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019. 2

[37] Jiankun Li, Peisen Wang, Pengfei Xiong, Tao Cai, Ziwei Yan, Lei Yang, Jiangyu Liu, Haoqiang Fan, and Shuaicheng Liu. Practical stereo matching via cascaded recurrent network with adaptive correlation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16263–16272, 2022. 5, 8

[38] Jiankun Li, Peisen Wang, Pengfei Xiong, Tao Cai, Ziwei Yan, Lei Yang, Jiangyu Liu, Haoqiang Fan, and Shuaicheng Liu. Practical stereo matching via cascaded recurrent network with adaptive correlation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16263–16272, 2022. 5

[39] Kunming Li, Yu Li, Shaodi You, and Nick Barnes. Photo-realistic simulation of road scene for data-driven methods in bad weather. *Proceedings of the IEEE International Conference on Computer Vision*, pages 491–500, 2017. 2

[40] Zhenyu Li, Zehui Chen, Xianming Liu, and Junjun Jiang. Depthformer: Exploiting long-range correlation and local information for accurate monocular depth estimation, 2022. 8

[41] Zhenyu Li, Xuyang Wang, Xianming Liu, and Junjun Jiang. Binsformer: Revisiting adaptive bins for monocular depth estimation, 2022. 8

[42] Huan Liu, Zijun Wu, Liangyan Li, Sadaf Salehkalaibar, Jun Chen, and Keyan Wang. Towards multi-domain single image dehazing via test-time training. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5821–5830, 2022. 2

[43] Yun-Fu Liu, Da-Wei Jaw, Shih-Chia Huang, and Jenq-Neng Hwang. Desnownet: Context-aware deep network for snow removal. *IEEE Transactions on Image Processing*, 27(6):3064–3073, 2018. 2

[44] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *International Conference on Learning Representations*, 2018. 5

[45] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2, 3, 4, 5, 6, 7

[46] Thomas Neff, Pascal Stadlbauer, Mathias Parger, Andreas Kurz, Joerg H Mueller, Chakravarty R Alla Chaitanya, Anton Kaplanyan, and Markus Steinberger. Donerf: Towards real-time rendering of compact neural radiance fields using depth oracle networks. In *Computer Graphics Forum*, volume 40, pages 45–59. Wiley Online Library, 2021. 2

[47] Julian Ost, Fahim Mannan, Nils Thuerey, Julian Knodt, and Felix Heide. Neural scene graphs for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2856–2865, 2021. 2, 6, 7

[48] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 5

[49] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8152–8160, 2019. 1, 2, 7

[50] Barbara Roessle, Jonathan T Barron, Ben Mildenhall, Pratul P Srinivasan, and Matthias Nießner. Dense depth priors for neural radiance fields from sparse input views. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12892–12901, 2022. 3, 5, 6, 7

[51] Viktor Rudnev, Mohamed Elgharib, William Smith, Lingjie Liu, Vladislav Golyanik, and Christian Theobalt. Neural radiance fields for outdoor scene relighting. *arXiv preprint arXiv:2112.05140*, 2021. 2

[52] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic foggy scene understanding with synthetic data. *International Journal of Computer Vision*, 126(9):973–992, 2018. 2, 3

[53] Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Semantic Nighttime Image Segmentation with Synthetic Stylized Data, Gradual Adaptation and Uncertainty-Aware Evaluation. *CoRR*, abs/1901.05946, 2019. 2

[54] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *CVPR*, 2019. 5

[55] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *CVPR*, 2020. 5

[56] Advaith Venkatramanan Sethuraman, Manikandasriram Srinivasan Ramanagopal, and Katherine A Skinner. Waternerf: Neural radiance fields for underwater scenes. *arXiv preprint arXiv:2209.13091*, 2022. 2

[57] Zheng Shi, Ethan Tseng, Mario Bijelic, Werner Ritter, and Felix Heide. Zeroscatter: Domain transfer for long distance imaging and vision through scattering media. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1, 2, 7

[58] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7495–7504, 2021. 2

[59] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *CVPR*, 2022. 6, 7

[60] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1

[61] Sabine Süsstrunk, Robert Buckley, and Steve Swen. Standard rgb color spaces. In *Proc. IS&T;/SID 7th Color Imaging Conference*, volume 7, pages 127–134, 1999. 4

[62] Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P Srinivasan, Jonathan T Barron, and Henrik Kretzschmar. Block-nerf: Scalable large scene neural view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8248–8258, 2022. 2

[63] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33:7537–7547, 2020. 4

[64] Ketan Tang, Jianchao Yang, and Jue Wang. Investigating haze-relevant features in a learning framework for image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2995–3000, 2014. 2, 4

[65] Maxime Tremblay, Shirsendu S. Halder, Raoul de Charette, and Jean-François Lalonde. Rain rendering for evaluating and improving robustness to bad weather. *International Journal of Computer Vision (IJCV)*, 126, 2021. 2

[66] Haithem Turki, Deva Ramanan, and Mahadev Satyanarayanan. Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12922–12931, 2022. 2

[67] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5481–5490. IEEE, 2022. 6, 7

[68] Alexander von Bernuth, Georg Volk, and Oliver Bringmann. Rendering Physically Correct Raindrops on Windshields for Robustness Verification of Camera-based Object Recognition. *IEEE Intelligent Vehicles Symposium (IV)*, pages 922–927, 2018. 2

[69] Can Wang, Menglei Chai, Mingming He, Dongdong Chen, and Jing Liao. Clip-nerf: Text-and-image driven manipulation of neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3835–3844, 2022. 2

[70] Qianyi Wu, Xian Liu, Yuedong Chen, Kejie Li, Chuanxia Zheng, Jianfei Cai, and Jianmin Zheng. Object-compositional neural implicit surfaces. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVII*, pages 197–213. Springer, 2022. 2

[71] Jiaqi Xu, Xiaowei Hu, Lei Zhu, Qi Dou, Jifeng Dai, Yu Qiao, and Pheng-Ann Heng. Video dehazing via a multi-range temporal alignment network with physical prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 7

[72] Tianhan Xu and Tatsuya Harada. Deforming radiance fields with cages. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIII*, pages 159–175. Springer, 2022. 2

[73] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2037–2046, June 2022. 1, 2, 7

[74] Yang Yang, Chaoyue Wang, Risheng Liu, Lin Zhang, Xiaojie Guo, and Dacheng Tao. Self-augmented unpaired image dehazing via density and depth decomposition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2037–2046, June 2022. 1

[75] Yu-Jie Yuan, Yang-Tian Sun, Yu-Kun Lai, Yuewen Ma, Rongfei Jia, and Lin Gao. Nerf-editing: geometry editing of neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18353–18364, 2022. 2

[76] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020. 2, 6, 7

[77] Karel Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. 1994. 1, 2