

RainyScape: Unsupervised Rainy Scene Reconstruction using Decoupled Neural Rendering

Paper ID: 3249

ABSTRACT

We propose RainyScape, an unsupervised framework for reconstructing clean scenes from a collection of multi-view rainy images. RainyScape consists of two main modules: a neural rendering module and a rain-prediction module that incorporates a predictor network and a learnable latent embedding that captures the rain characteristics of the scene. Specifically, based on the spectral bias property of neural networks, we first optimize the neural rendering pipeline to obtain a low-frequency scene representation. Subsequently, we jointly optimize the two modules, driven by the proposed adaptive direction-sensitive gradient-based reconstruction loss, which encourages the network to distinguish between scene details and rain streaks, facilitating the propagation of gradients to the relevant components. Extensive experiments on both the classic neural radiance field and the recently proposed 3D Gaussian splatting demonstrate the superiority of our method in effectively eliminating rain streaks and rendering clean images, achieving state-of-the-art performance. The constructed high-quality dataset and source code will be publicly available.

CCS CONCEPTS

• **Computing methodologies** → **Reconstruction**; *3D imaging*; Computational photography.

KEYWORDS

Rainy scene reconstruction, Neural rendering, Unsupervised loss

1 INTRODUCTION

Neural Radiance Field (NeRF) [29] has emerged as a groundbreaking technique for novel view synthesis by learning a continuous, volumetric representation of the scene through differentiable volume rendering. NeRF’s ability to create highly realistic and consistent novel views with fine details has led to its widespread adoption in various applications, such as 3D reconstruction [40], surface reconstruction [3], 3D object editing [36], and large-scale scene reconstruction [53].

However, when the input images are degraded by various factors such as blur, noise, or rain, the rendering results inevitably exhibit obvious artifacts. To address this issue, recent works have proposed a range of task-specific solutions. For instance, Ma et al. [27] achieved clear scene reconstruction from images affected by camera

motion blur or defocus blur by learning ray fusion within the blur kernel. Huang et al. [16] tackled the problem of input images with different dynamic ranges by adjusting the ray intensity using a camera response function and obtaining the final color with a learnable tone mapper. Chen et al. [6] extended the atmospheric scattering model into the volume rendering process to fit hazy images using two neural rendering processes. In contrast to these methods, our approach specifically targets the rainy scene reconstruction task, which can be seamlessly integrated with our observation of the neural rendering prior, as illustrated in Fig. 2. Moreover, the sparse and intermittent nature of rain precipitation in 3D space makes it challenging to represent through an additional neural rendering field. Furthermore, our proposed framework can be readily adapted to work with various rendering techniques, demonstrating its versatility and flexibility.

In this paper, we propose RainyScape, a decoupled neural rendering framework that is capable of reconstructing a rain-free scene from rainy images in an unsupervised fashion. First, we obtain a low-frequency representation of the scene through a selected radiance field rendering process (e.g., NeRF [29] or 3D Gaussian Splatting [19]), where the remaining high-frequency scene details and rain information are coupled together. We then characterize the rain in the scene using learnable rain embeddings composed of rainy scene state vectors, viewpoint state vectors, and camera parameters, and predict rain maps through a multilayer perceptron (MLP) and a CNN predictor. To decouple the high-frequency scene details from the rain streak, we propose an adaptive angle estimate strategy that significantly improves the distinguishability of rain by considering the directions along and perpendicular to the rain streak. We use the obtained angle information to design a gradient rotation loss. Combining our proposed network framework and unsupervised loss, we employ an alternating optimization strategy to update network parameters and rain embeddings. Additionally, to address the lack of multi-view rainy scene datasets, we render 10 sets of scenes using Maya [1], resulting in more consistent and realistic rain trails compared to data simulated by simple methods.

In summary, the main contributions of this paper lie in:

- we explore priors in neural rendering processes and propose a general rainy scene reconstruction framework;
- we represent the rain in the scene using rain embeddings and use a predictor to predict rain streaks;
- we introduce an adaptive scene rain streak angle estimate strategy and a corresponding gradient rotation loss for decoupling scene high-frequency details and rain streaks; and
- we construct a multi-view rainy scene dataset for more realistic and consistent rain streaks.

Permission to make digital or hard copies of all or part of this work for personal or professional use, not for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM MM, 2024, Melbourne, Australia
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
<https://doi.org/10.1145/nnnnnnn.nnnnnn>

Unpublished working draft. Not for distribution.

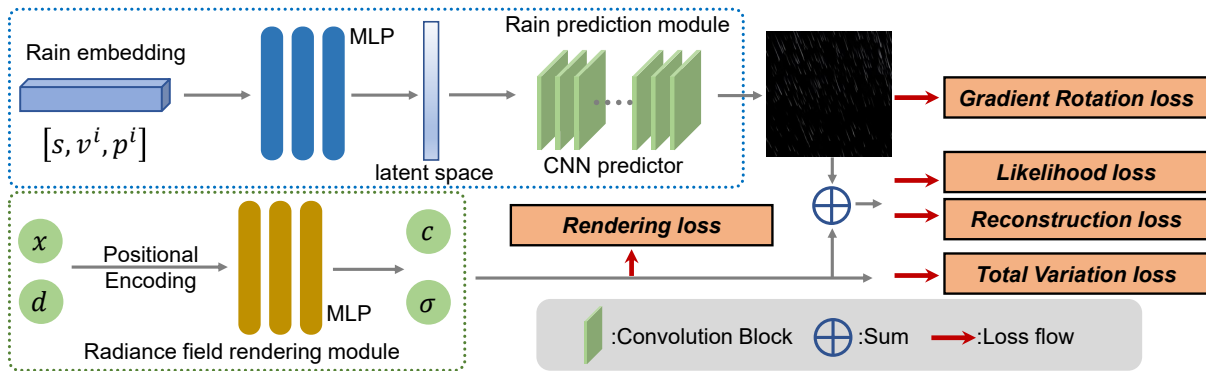


Figure 1: Overview of the proposed RainyScape framework, which can reconstruct a rain-free scene from a set of multi-view rainy images in an unsupervised fashion. Based on the NeRF architecture, the rendering module takes ray positions and view directions as input to estimate color and density values. Rain characteristics are modeled using scene state vectors s , viewpoint state vectors v^i , and camera parameters p^i . The combined rain embedding is processed through an MLP to obtain latent space representations, which are then fed into a CNN predictor to produce a rain map. The framework is trained using unsupervised losses that facilitate the decoupling of high-frequency scene details and rain streaks, yielding a rain-free neural radiance field.

2 RELATED WORK

2.1 Neural Rendering

NeRF [29] revolutionized viewpoint generation by introducing differentiable volume rendering and learnable radiance fields, enabling joint optimization of geometry and appearance for static 3D scenes from posed RGB images. Subsequent works explore various designs to improve NeRF’s training or inference speed, such as MLP capacity [34], space discretization [10]. Recently, Kerbl et al. [19] extended the NeRF to explicit GPU-friendly 3D Gaussians Splatting (3DGS) and replaced the neural rendering, achieving real-time rendering of radiance fields. Other works focus on enhancing rendering quality and addressing challenging input scenarios. For instance, PixelNeRF [48] enables few-shot reconstruction, Mip-NeRF [2] improves rendering under multi-scale inputs, Deblur-NeRF [27] handles blurry images using canonical kernels, and NaN [31] leverages inter-view and spatial information to deal with noisy data. HDR-NeRF and RawNeRF [16, 28] utilize camera response functions and raw images to process low dynamic range inputs. In contrast, our work tackles the novel task of rainy scene reconstruction using decoupled neural rendering. We discover that the neural rendering prior is well-suited for rainy scene reconstruction and propose a general framework based on this insight, setting our approach apart from existing NeRF-based methods.

2.2 Image/Video Deraining

Single-image rain removal is an ill-posed problem that aims to decompose a rainy image into a clean background and a rain streak layer. Traditional methods rely on prior characteristics of rain, such as photometric properties [13], morphological component analysis [18], non-local means filtering [20], and layer priors [24] to tackle this challenge. Other approaches employ optimization techniques, including sparse coding [26], Gaussian mixture models [46], and low-rank models [8], to separate rain streaks from the background. With the advent of deep learning, numerous learning-based methods [5, 7, 11, 15, 23, 30, 32, 38, 39, 41, 49, 51] have emerged,

significantly advancing the state-of-the-art in single image rain removal performance.

Video provides additional temporal information that can be exploited for rain removal. Garg et al. [12] pioneered the task using photometric properties. Chen et al. [4] employed superpixel segmentation and a robust deep CNN for restoration. Li et al. [22] proposed a multi-scale convolutional sparse coding model to capture repetitive local patterns and rain streaks of different scales. Yang et al. introduced a two-stage recurrent framework [44] and extended it to a self-learned approach [45] leveraging temporal correlation and consistency. Yue et al [50] proposed a semi-supervised framework with a dynamical rain generator. Yan et al. developed SLDNet+ [43], an augmented self-learned deraining network utilizing temporal information and rain-related priors, and combined single-image and multi-frame modules for raindrop removal. Zhang et al. [52] proposed ESTINet, an efficient end-to-end framework based on deep residual networks and convolutional LSTMs, to capture spatial features and temporal correlations among successive frames. Xiao et al. [25] introduced a teacher-student framework for adaptive nighttime video deraining.

In addition to video-based methods, some works tackle structured multi-view image rain removal. Ding et al. [9] proposed a GAN-based architecture utilizing depth information to remove rain streaks from 3D EPIs of rainy light field images (LFIs). Yan et al. [42] employed 4D convolutions and multi-scale Gaussian processes for LFI rain removal.

2.3 Preliminary

In this work, we employ the classic Neural Radiance Fields (NeRF) [29] as the neural rendering method to describe our work more concisely and elegantly. NeRF is a novel view synthesis method that learns a continuous, volumetric representation of a scene through differentiable volume rendering. NeRF employs a multilayer perceptron (MLP) network F_{Θ} to parameterize a 5D input (3D position $\mathbf{x} = (x, y, z)$ and 2D viewing direction $\mathbf{d} = (\theta, \phi)$) into a density σ



Figure 2: Illustration of neural rendering prior: From low-frequency scene representation to high-frequency detail restoration and rain preservation. The numbers represent the training epoch, the PSNR (dB) of the rendered image compared with rain-free ground truth images, and the PSNR (dB) of the rendered image compared with the rainy image.

and color $\mathbf{c} = (r, g, b)$:

$$F_{\Theta} : (\mathbf{x}, \mathbf{d}) \rightarrow (\sigma, \mathbf{c}) \quad (1)$$

To render a novel view, NeRF casts a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ for each pixel, where \mathbf{o} is the camera origin and \mathbf{d} is the ray direction. The expected color $C(\mathbf{r})$ of the ray is computed by integrating the radiance along the ray:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})dt, \quad (2)$$

where $T(t) = \exp(-\int_{t_n}^t \sigma(\mathbf{r}(s))ds)$ is the accumulated transmittance along the ray, and $[t_n, t_f]$ is the near and far bounds of the ray. In practice, the continuous integral is approximated using numerical quadrature:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^N T_i(1 - \exp(-\sigma_i\delta_i))\mathbf{c}_i, \quad (3)$$

where $T_i = \exp(-\sum_{j=1}^{i-1} \sigma_j\delta_j)$, $\delta_i = t_{i+1} - t_i$ is the distance between adjacent samples, and N is the number of samples along the ray.

3 PROPOSED METHOD

As aforementioned, given a set of multi-view rainy images, we aim to reconstruct a rain-free scene in an unsupervised manner. As shown in Fig. 1, the proposed framework, named RainyScape, consists of a neural rendering module based on the NeRF architecture and a rain prediction module built upon MLP and CNN layers. To be specific, the neural rendering module takes ray positions and view directions as input to estimate color and density values. The rain prediction module explicitly models the rain characteristics, which consumes a learnable rain embedding composed of scene state vectors and viewpoint state vectors, as well as provided camera parameters, to predict a rain map. During training, we alternately optimize the entire framework and use unsupervised losses to encourage the decoupling of high-frequency scene details and rain streaks. This enables the reconstruction of a rain-free radiance field, which can be directly rendered during inference to obtain sharp, rain-free views. Moreover, our framework can be adapted to other rendering pipelines, such as the recent 3D Gaussian Splatting (3DGS), as experimentally demonstrated in Section 5.

Remark. Our RainyScape differs from existing image/video-based deraining methods in several key aspects. First, those methods rely on image space operations and do not fully exploit the scene’s 3D geometry. Second, most of them require ground-truth rain-free data as supervision. Third, those methods often struggle to

handle multi-view inputs with large baselines. Furthermore, by reconstructing a rain-free radiance field, our RainyScape enables the rendering of novel views, which is not applicable to existing ones.

3.1 Neural Rendering Prior for Deraining

The task of deraining aims to decompose a rainy image $\mathbf{I} \in \mathbb{R}^{h \times w \times 3}$ into a rain-free background scene $\mathbf{B} \in \mathbb{R}^{h \times w \times 3}$ and a rain layer $\mathbf{R} \in \mathbb{R}^{h \times w \times 3}$ [35, 37], which can be formulated as

$$\mathbf{I} = \mathbf{B} + \mathbf{R}. \quad (4)$$

Initially, we leverage the neural rendering prior, which is fundamentally introduced by the spectral bias property of neural networks [33], to obtain a low-frequency scene representation $\mathbf{B}_l \approx \mathbf{B}$ during the warm-up stage, as illustrated in Fig. 2. Specifically, at this stage, the rainy image \mathbf{I} can be decomposed into the low-frequency scene content \mathbf{B}_l and the high-frequency information \mathbf{I}_h , as shown in the residual map in Fig. 3, which includes both rain streaks and useful scene details:

$$\mathbf{I} = \mathbf{B}_l + \mathbf{I}_h. \quad (5)$$

To obtain a more detailed background \mathbf{B} from the initial low-frequency representation \mathbf{B}_l , we need to further decouple the high-frequency information \mathbf{I}_h into rain streaks and detailed scene components. To achieve this goal, we next introduce a rain prediction module that receives and reflects the rain characteristics from the rainy scene.

3.2 Rain Prediction Module

To effectively model the rain characteristics and decouple the high-frequency information \mathbf{I}_h into rain streaks \mathbf{R} and detailed scene components \mathbf{B}_h , we introduce a rain prediction module E_{Φ} . This module consists of a three-layer MLP to extract a latent space representation of the rain characteristics and a six-layer CNN to process and upsample feature maps to obtain a rain map. It takes learnable rain embeddings as input and predicts a rain map $\mathbf{R}_e \in \mathbb{R}^{h \times w \times 3}$.

The rain embeddings comprise three parts: learnable scene state vectors $\mathbf{s} \in \mathbb{R}^{128}$, learnable viewpoint state vectors $\mathbf{v} \in \mathbb{R}^{64 \times n}$, and fixed camera parameters $\mathbf{p} \in \mathbb{R}^{16 \times n}$, where n is the number of input views. For a specific view i , the predicted rain streaks can be expressed as

$$\mathbf{R}_e^i = E_{\Phi}(\mathbf{s}, \mathbf{v}^i, \mathbf{p}^i). \quad (6)$$

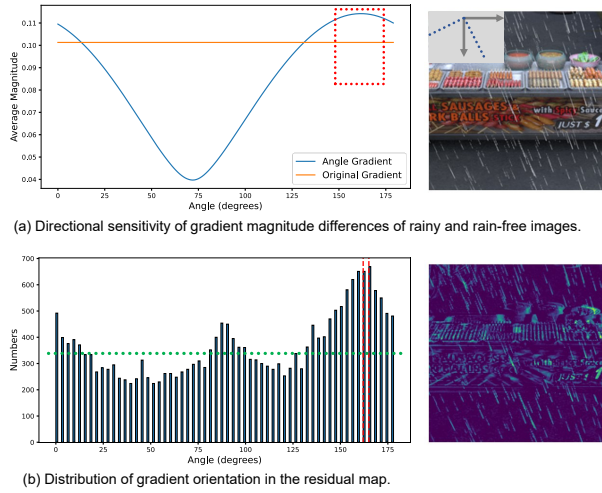


Figure 3: Leveraging directional sensitivity and gradient orientation for unsupervised rainy scene reconstruction. (a) Directional sensitivity of gradient magnitude differences between rainy and rain-free images. Gradients perpendicular to the rain direction exhibit higher discriminative power compared to those along the rain direction. (b) Distribution of gradient orientation in the residual map $\mathbf{I} - \mathbf{B}_I$. The residual map contains gradients in all directions (green dashed lines), with a dominant orientation perpendicular to the rain direction (red dashed lines) due to the presence of rain streaks.

These parameters are designed to capture and represent essential information about the rain in the scene that influences the appearance of rain streaks. The rainy scene state vectors \mathbf{s} represent global rainy characteristics, such as the scale and direction of the rain. The viewpoint state vectors \mathbf{v} record view-specific characteristics, allowing for subtle changes in the state of rain between viewpoints. The camera parameters \mathbf{p} represent the viewing direction and position, affecting the rain streaks' orientation and perspective.

3.3 Unsupervised Loss Function

In addition to the proposed rain prediction model, we propose a novel unsupervised loss function to help the network effectively decouple high-frequency details and rain streaks. A key component of our loss function is the adaptive gradient rotation loss, which plays a crucial role in distinguishing rain streaks from background textures.

Adaptive Gradient Rotation Loss. As illustrated in Fig. 3 (a), angle information can be used to differentiate rain streaks from background images. Unlike existing works, such as [17] that shifts the image through several preset angles and [49] that requires pre-training an angle estimation network, we propose an adaptive rain streak angle estimation strategy based on several experimental observations:

- The residual map $\mathbf{I} - \mathbf{B}_I$ contains high-frequency details and rain, with the angles of high-frequency details in the scene being relatively evenly distributed.

- The gradient direction of the rain is dominant after suppressing the minimum gradient.
- For local image patches, there is a high probability of rain occurring in only one direction.

Based on these observations, we calculate the orientation of the residual image gradients and construct a histogram to identify the dominant rain streak directions. Considering that the discrimination does not change significantly within a small angle range, we discretize the orientation range into 60 bins, each spanning an angle of 3 degrees for robustness. Additionally, for the special case of rain with multiple angles, we can obtain multiple angles by adjusting the number of top angles considered.

After adaptively obtaining the rain angle, for each dominant direction θ , the adaptive gradient rotation loss is given by

$$\mathcal{L}_{agr} = \frac{1}{K} \sum_{j=1}^K (|\nabla_{\theta_k + \pi/2} \mathbf{R}| - |\nabla_{\theta_k} \mathbf{R}| + |\nabla_{\theta_k} \mathbf{B}| + |\nabla_{\theta_k} (\mathbf{I} - \mathbf{R})|), \quad (7)$$

where K is the number of dominant directions, typically set to 1, and ∇_{θ} denotes the gradient operator along the direction θ .

Likelihood Loss. The likelihood loss measures the discrepancy between the reconstructed image $\mathbf{B} + \mathbf{R}$ and the input rainy image, promoting stable network performance:

$$\mathcal{L}_{ll} = \frac{1}{(\sigma^2 + \epsilon)} \cdot \|\mathbf{I} - \mathbf{B} - \mathbf{R}\|_2^2, \quad (8)$$

where σ is the standard deviation of the residual, and ϵ is a small constant to prevent division by zero.

Reconstruction Loss. The reconstruction loss further emphasizes the consistency between the reconstructed image and the input rainy image:

$$\mathcal{L}_{rec} = \|\mathbf{I} - \mathbf{B} - \mathbf{R}\|_2^2. \quad (9)$$

Total Variation Loss. The total variation loss encourages the rendered background image to be smooth, reducing artifacts and noise:

$$\mathcal{L}_{tv} = |\nabla_x \mathbf{B}|_1 + |\nabla_y \mathbf{B}|_1, \quad (10)$$

where ∇_x and ∇_y denote the horizontal and vertical gradient operators, respectively.

The overall loss function is written as

$$\mathcal{L} = \lambda_1 \mathcal{L}_{ll} + \lambda_2 \mathcal{L}_{rec} + \lambda_3 \mathcal{L}_{tv} + \lambda_4 \mathcal{L}_{agr}, \quad (11)$$

where λ_1 , λ_2 , λ_3 , and λ_4 are the weight coefficients for the corresponding loss terms.

3.4 Optimization Process

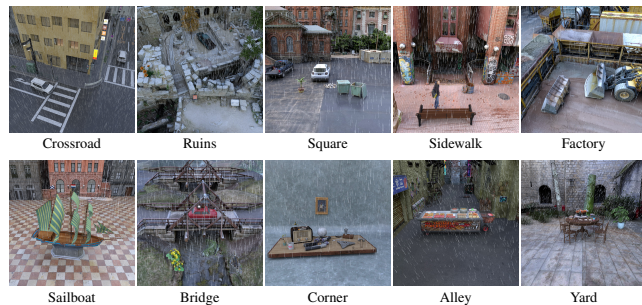
We propose an alternating optimization approach to train our framework effectively, as outlined in Algorithm 1. The optimization process consists of two main stages: network update and latent variable update. During the network update stage, we predict rain maps using the predictor E_{Φ} , render background rays using the NeRF network F_{Θ} , and update the network parameters Θ and Φ based on the defined loss functions Eq. (11). In the latent variable update stage, we freeze the networks and update the latent variables \mathbf{s} and \mathbf{v} using Langevin Monte Carlo [14], which introduces random perturbations to explore the latent space and discover better representations for fitting the input rainy image \mathbf{I} . The optimization process iterates until convergence, alternating between the two stages, with

Table 1: Performance comparison of different methods on various data. The best performance is shown in bold, and the second-best performance is underlined. The Supervised indicates whether utilizes additional supervision information.

Method	Supervised	Crossroad			Square			Sailboat			Yard			Average		
		PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
DRSF [7]+NeRF	✓	29.76	0.823	0.272	27.43	0.789	0.266	26.82	0.767	0.281	29.63	0.805	0.222	28.41	0.796	0.260
EIST [52]+NeRF	✓	29.54	0.819	0.256	27.35	0.789	0.269	26.42	0.753	0.307	29.37	0.796	0.231	28.17	0.789	0.266
NeRF	✗	27.54	0.836	0.146	26.44	<u>0.824</u>	0.136	26.34	<u>0.822</u>	0.067	27.39	0.796	0.099	26.93	<u>0.820</u>	0.112
Ours	✗	31.08	0.876	<u>0.148</u>	29.58	0.870	0.135	29.59	0.890	<u>0.069</u>	29.96	0.861	<u>0.104</u>	30.05	0.874	<u>0.114</u>

Algorithm 1 Alternating Optimization for RainyScape

- 1: **Input:** rainy image I , camera poses \mathbf{p} .
- 2: **Output:** optimized NeRF network F_{Θ} , predictor E_{Φ} , and latent variables \mathbf{s} and \mathbf{v} .
- 3: Initialize Θ , Φ , \mathbf{s} , and \mathbf{v} .
- 4: **while** not converged **do**
- 5: **if** $epoch < 50\%$ total epoch **then**
- 6: Warm-up NeRF network F_{Θ} .
- 7: **else**
- 8: **Network Update:** Predict rain maps using E_{Φ} . Render background rays using F_{Θ} , and Update network parameter Θ and Φ via Eq. 11.
- 9: **Latent Variable Update:** Freeze F_{Θ} and E_{Φ} . Update \mathbf{s} , and \mathbf{v} using Langevin Monte Carlo to fit I via Eq. 11.
- 10: **end if**
- 11: **end while**

**Figure 4: Scenes and names of the proposed dataset.**

a warm-up stage to initialize the NeRF network before introducing the rain prediction module. By leveraging this approach, our framework learns to separately represent the background scene and the rain layer, progressively refining the deraining results without relying on ground-truth supervision.

4 EXPERIMENTS

4.1 Proposed Dataset

We construct a comprehensive and diverse multi-view rainy image dataset using *Maya* [1]. As illustrated in Fig. 4, our dataset consists of 10 scenes, each captured from 50 viewpoints with 1024×1024 resolution. To ensure the view consistency of raindrops, we construct 3D models of raindrops, distribute them randomly within the scene, and render them together with the scene objects. This approach enables the generation of interaction effects between raindrops and light rays, generating more realistic rainy scenes compared to

simulation methods that directly add rain streaks to 2D images. The dataset encompasses a wide range of rain densities, directions, and streak orientations to accurately represent real-world variations. The scenes are composed of outdoor objects such as cars, bridges, roads, buildings, and boats to facilitate generalization to unseen rainy environments. The 50 viewpoints are strategically distributed on a spherical surface to maximize the overlap rate of scene content captured by different cameras, providing diverse information for rendering. The dataset includes rainy and ground truth images, depth maps, and camera parameters, making it suitable for a wide range of viewpoint synthesis tasks. *We will make the dataset publicly available.*

4.2 Implementation Details

Our rain prediction module consists of a 3-layer MLP with a feature dimension of 128 channels, which generates a 1024-dimensional latent space representation. This representation is then reshaped into an image and processed using a 6-layer convolutional network, followed by super-resolution to obtain the required 64×64 random patches. For the neural rendering component, we utilize the PyTorch re-implementation NeRF [47] with a batch size of 4096 rays, each ray undergoing 64 coarse samplings and 64 fine samplings. To further validate our approach, we also conduct experiments using the official implementation of 3D Gaussian Splatting (3DGS) [19] with its default configuration. During model optimization, we employ the Adam optimizer [21] with its default settings. For NeRF, we use the default learning rate update strategy, with an initial learning rate of 5×10^{-4} . After the warm-up period, the learning rate is set to 1×10^{-6} . The learning rates for the MLP and CNN part in the rain prediction module are set to 1×10^{-3} and 1×10^{-4} , respectively. For latent variable updates, we perform 5 updates per epoch, with the first two updates using Langevin Monte Carlo to introduce random perturbations. The loss coefficients λ_1 , λ_2 , λ_3 , and λ_4 are set to 0.1, 500, 0.5, and 1, respectively. We optimize a single model for 18K iterations on a single NVIDIA A6000 GPU.

4.3 Results

We set up two comparison baselines that leverage supervised information for rain removal. First, we fine-tune two state-of-the-art rain removal methods, namely the single-image rain removal method DRSformer [7] and the video-based rain removal method ESTINet [52], using six randomly selected scenes (Ruins, Sidewalk, Factory, Bridge, Corner, and Alley). These fine-tuned models are then employed as pre-processing steps for the input images. Subsequently, we train NeRF on the pre-processed images and refer to these two

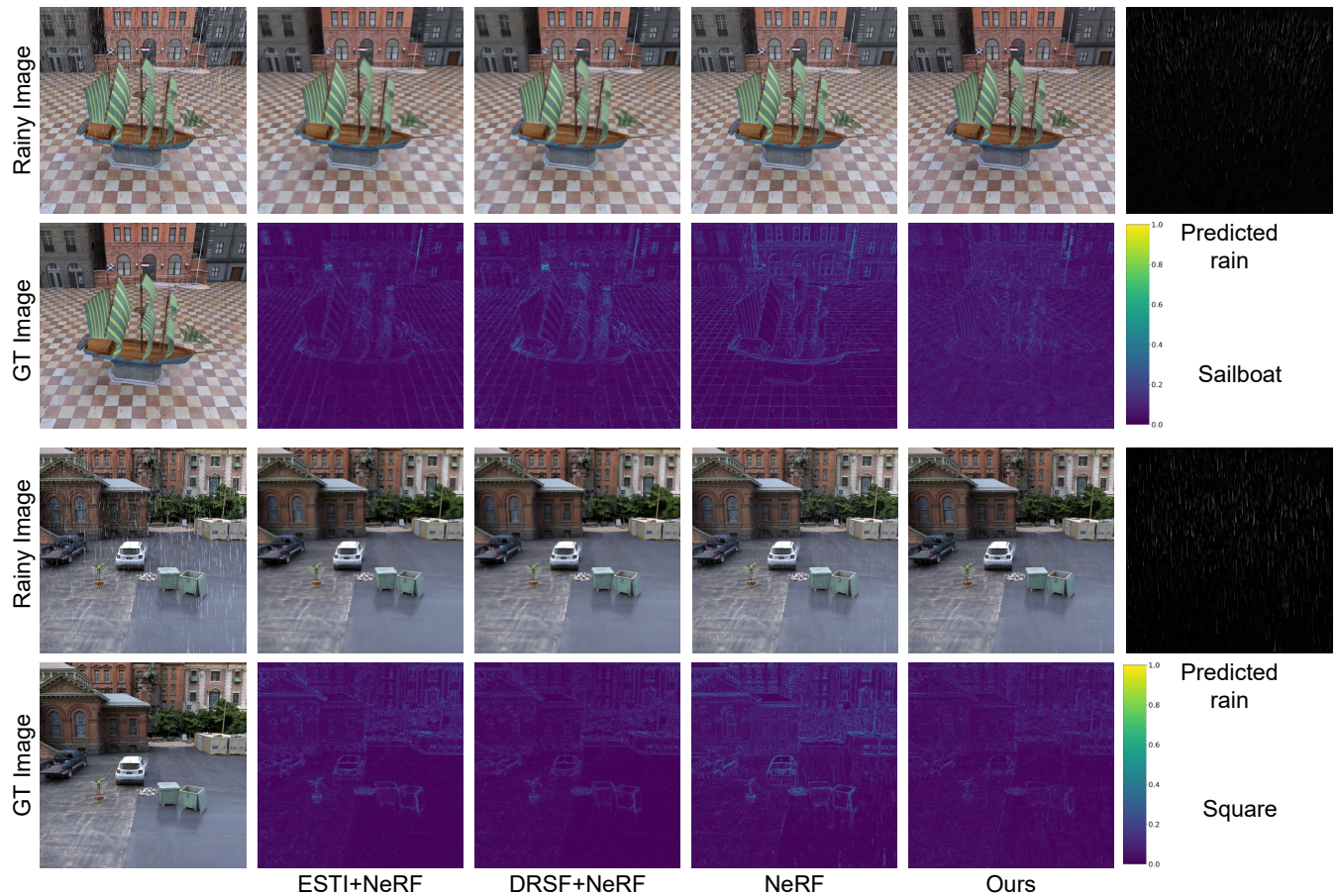


Figure 5: Visual comparison of different methods on rainy scenes. Each scene shows rainy images, ground truth, rendered results from baseline methods and our approach, error maps, and rain streak images predicted by our method.

approaches as DRSF+NeRF and ESTI+NeRF, respectively. Additionally, we include a baseline that directly applies the original NeRF [29] to the rainy data without any pre-processing or modifications.

We employ widely-used metrics to quantitatively evaluate the performance of our method and the baselines. Peak Signal-to-Noise Ratio (PSNR) measures the image quality and rainy scene reconstruction performance, with higher values indicating better results. Structural Similarity Index Measure (SSIM) assesses quality via structural information, luminance, and contrast, with higher values indicating better results. Perceptual similarity (LPIPS) measures the similarity in feature space, with lower values indicating better results.

Quantitative results. Table 1 showcases the performance comparison of our proposed method against three baselines (DRSF+NeRF, EIST+NeRF, and NeRF) on four scenes (Crossroad, Square, Sailboat, and Yard). The "Supervised" column in the table indicates whether the method utilizes supervision information. Our method achieves the best performance on the majority of the scenes and metrics (highlighted in **bold**), outperforming the baselines in most cases. On average, our approach obtains a PSNR of 29.88 dB, surpassing the second-best method, DRSF+NeRF, by 1.47 dB and the original

NeRF by 2.95 dB. We also achieve the highest average SSIM and second-best average LPIPS scores, indicating superior structural preservation and perceptual quality. It is worth noting that our method achieves state-of-the-art performance without relying on any supervision information. These quantitative results validate the effectiveness of our novel contributions and highlight its potential to render high-quality clear images across diverse rainy scenes.

Qualitative results. Fig. 5 presents a visual comparison of the rendering results obtained by our method and the baseline approaches. Our approach effectively renders rain-free images while accurately capturing and separating the rain streaks from the scene content. The error maps demonstrate our method's superiority in recovering high-frequency details and preserving textures and edges compared to the baselines. By explicitly modeling and decoupling the rain streaks from the high-frequency content, our framework can effectively transfer useful scene details to the radiance field, resulting in improved restoration quality. In contrast, the baseline methods often struggle to recover fine details, introducing artifacts or losing important texture information. Since the rain streaks are relatively small compared to the scene, please zoom in for better observation of the details in all figures.

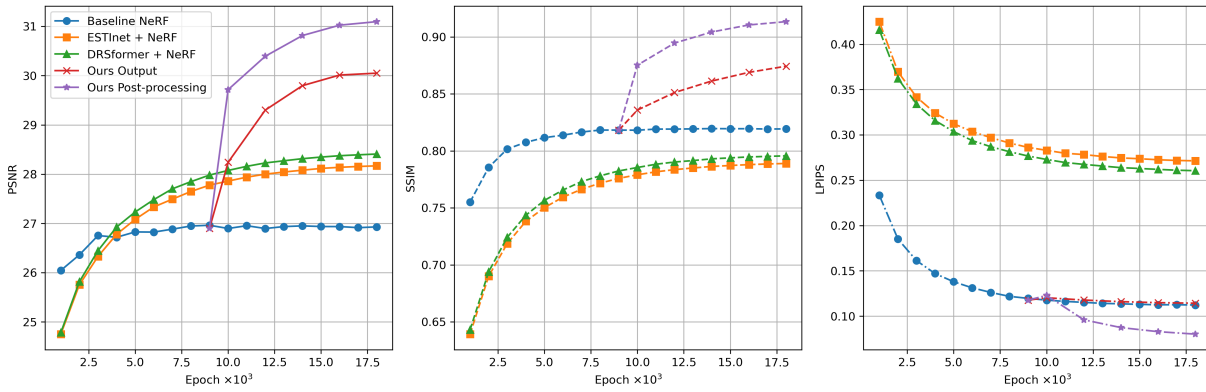


Figure 6: Average performance of different methods over training epochs on four scenes. Left: PSNR, Middle: SSIM, Right: LPIPS. Our post-processing step consistently improves the deraining quality compared to our base output.

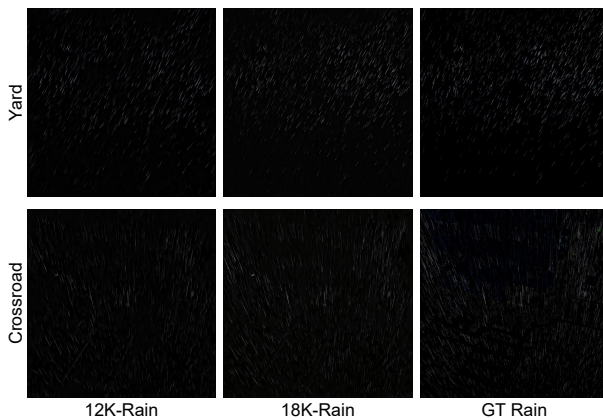


Figure 7: Evolution of predicted rain streaks during training and comparison with ground truth (GT) rain.

Post-processing with predicted rain. The accuracy of the predicted rain streak images, as demonstrated in Fig. 5 and Fig. 7, enables further enhancement of the deraining performance through post-processing. When the rainy image is provided, we can apply a simple thresholding method to convert the rain streak image into a binary mask, which allows us to identify and selectively fuse regions unaffected by rain with our rendered image. This selective fusion approach leads to an improvement in the overall deraining performance. Fig. 6 illustrates the performance progression of different methods, including our post-processing step, during training. By combining the information from the rain-free regions with our RainyScape rendering results, we achieve a further performance improvement of approximately 1 dB.

4.4 Ablation Study

Rain Prediction. To demonstrate the effectiveness of our rain prediction module, we first visualize the evolution of the predicted rain streaks during training in Fig. 7. As the training progresses, the predicted rain streaks gradually resemble the real rain streaks more closely, indicating that our model effectively learns to capture the characteristics of scene rain.

Furthermore, we investigate the impact of the rain embedding on the predicted rain streaks. In Fig. 8, we present the results of

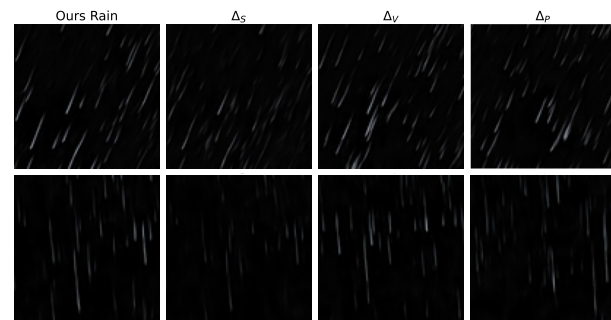


Figure 8: Impact of rain embedding changes (Δ_s , Δ_v and Δ_p) on predicted rain streaks, with 100×100 patch size.

adding random Gaussian noise with a variance of 0.5 to the scene state vector \mathbf{s} and viewpoint state vector \mathbf{v} , denoted as Δ_s and Δ_v , respectively. The results demonstrate that the added noise influences the intensity and density distribution of the predicted rain streaks, suggesting that the rain embedding plays a crucial role in controlling the appearance of the predicted rain. In addition, we show the effect of changing the camera parameters, denoted as Δ_p . The predicted rain adapts to the changes in camera parameters, demonstrating the ability to predict rain streaks that are consistent with the scene information and camera setup.

Loss configuration. To investigate the contribution of each term in our loss function (Eq. (11)), we conduct an ablation study by removing the reconstruction loss \mathcal{L}_{rec} , total variation loss \mathcal{L}_{tv} , and adaptive gradient rotation loss \mathcal{L}_{adg} terms individually. The likelihood loss \mathcal{L}_{ll} is kept in all configurations to ensure stable training. Additionally, we explore the impact of varying the number of bins used in the adaptive gradient rotation loss, considering 30, 60 (default), and 90 bins. All the results are shown in Table 2. Notably, when the \mathcal{L}_{adg} is not used, the network fails to predict visually realistic rain streaks, emphasizing its critical role in capturing the directional characteristics of rain.

Framework Effectiveness. We compare our framework with two alternative approaches to demonstrate its effectiveness. The first approach, Two-NeRF, uses separate neural rendering fields for rain and background, similar to Dehaze-NeRF [6]. However, as shown in Fig. 10, Two-NeRF fails to effectively decouple rain

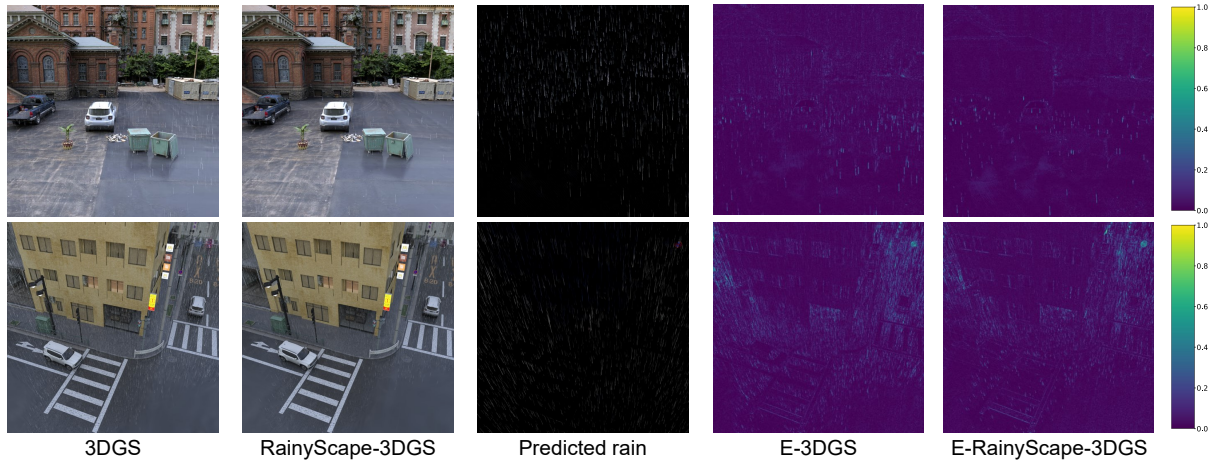


Figure 9: Visual results of the extension of our RainyScape to 3DGS. The error maps are denoted as E-*

Table 2: Ablation study of loss configuration on the Yard data.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
w/o \mathcal{L}_{rec}	12.52	0.428	0.467
w/o \mathcal{L}_{tv}	29.51	0.850	0.108
w/o \mathcal{L}_{agr}	29.28	0.856	0.105
30 bins	29.63	0.859	0.106
90 bins	29.60	0.858	0.106
Ours	29.96	0.861	0.104

Table 3: Average performances of 3DGS extension.

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
3DGS	32.41	0.907	0.219
RainyScape-3DGS	35.66	0.947	0.067
RainyScape-3DGS (Post-Process)	36.48	0.952	0.056

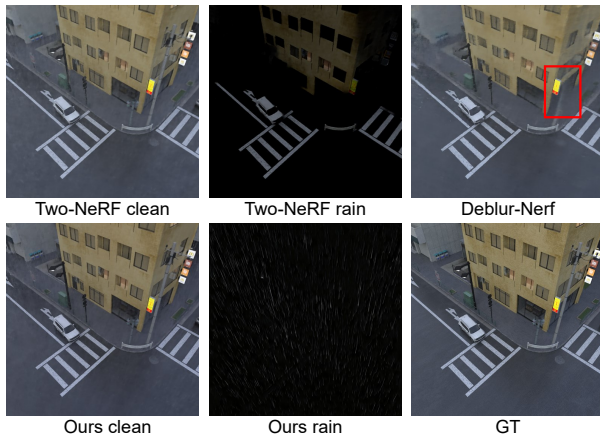


Figure 10: Limitations of alternative approaches. Two-NeRF fails to decouple rain and background. Deblur-NeRF damages scene structure (red box).

from the background, while our method accurately distinguishes between the two components. We also evaluate Deblur-NeRF [27], which adjusts and fuses rays to handle motion blur. As highlighted in Fig. 10, Deblur-NeRF severely damages the scene structure, erroneously removing objects like utility poles.

5 EXTENSION TO 3D GAUSSIAN SPLATTING

3D Gaussian splatting (3DGS) [19] has emerged as a promising technique in the field of radiance field rendering, offering high

efficiency and impressive visual quality. In this section, we explore the extension of our RainyScape framework to incorporate 3DGS as the rendering module. By leveraging the capabilities of 3DGS, we aim to enhance the deraining performance while maintaining the computational efficiency and visual fidelity of the rendered results.

To extend our framework to 3DGS, we replace the NeRF rendering component in the pipeline with the 3DGS architecture, getting RainyScape-3DGS, while the rain prediction modules remain unchanged. To ensure optimal performance, the warm-up stage is set to 4K iterations. Subsequently, joint training of the entire framework is performed using the default configurations and loss functions. We evaluate the effectiveness of RainyScape-3DGS using the same datasets. Table 3 presents the quantitative results, proving that our RainyScape framework can significantly improve the performance of 3DGS in rainy scenes. Fig. 9 provides visual comparisons, showcasing the ability of our framework to suppress the rendered rain in 3DGS and more visually appealing rain-free images.

6 CONCLUSION

In this paper, we introduced RainyScape, a novel unsupervised framework for reconstructing scenes from multi-view rainy images using decoupled neural rendering. Our approach tackles the challenges of rain streak removal in the rendering process by integrating a low-frequency scene representation, a rain prediction module with a learnable rain embedding, and an unsupervised rainy scene reconstruction loss that incorporates an adaptive gradient rotation loss. Extensive experiments on both classic NeRF and state-of-the-art 3DGS demonstrate the effectiveness and versatility of RainyScape in generating clean, visually appealing images with sharp details and accurate scene geometry.

REFERENCES

- [1] Autodesk, INC. [n. d.]. *Maya*. <https://autodesk.com/maya>
- [2] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. 2021. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5855–5864.
- [3] Jia-Wang Bian, Wenjing Bian, Victor Adrian Prisacariu, and Philip Torr. 2024. PoRF: Pose Residual Field for Accurate Neural Surface Reconstruction. In *International Conference on Learning Representations (ICLR)*.
- [4] Jie Chen, Cheen-Hau Tan, Junhui Hou, Lap-Pui Chau, and He Li. 2018. Robust video content alignment and compensation for rain removal in a cnn framework. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6286–6295.
- [5] Sixiang Chen, Tian Ye, Jinbin Bai, Erkang Chen, Jun Shi, and Lei Zhu. 2023. Sparse sampling transformer with uncertainty-driven ranking for unified removal of raindrops and rain streaks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 13106–13117.
- [6] W. Chen, W. Yifan, S. Kuo, and G. Wetzstein. 2024. DehazeNeRF: Multiple Image Haze Removal and 3D Shape Reconstruction using Neural Radiance Fields. In *3DV*.
- [7] Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan. 2023. Learning a sparse transformer network for effective image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5896–5905.
- [8] Yi-Lei Chen and Chiou-Ting Hsu. 2013. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *Proceedings of the IEEE international conference on computer vision*. 1968–1975.
- [9] Yuyang Ding, Mingyue Li, Tao Yan, Fan Zhang, Yuan Liu, and Rynson WH Lau. 2021. Rain streak removal from light field images. *IEEE Transactions on Circuits and Systems for Video Technology* 32, 2 (2021), 467–482.
- [10] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. 2022. Plenoxels: Radiance Fields Without Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5501–5510.
- [11] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. 2017. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3855–3863.
- [12] Kshitiz Garg and Shree K Nayar. 2004. Detection and removal of rain from videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vol. 1. 1–1.
- [13] Kshitiz Garg and Shree K Nayar. 2005. When does a camera see rain?. In *IEEE International Conference on Computer Vision*, Vol. 2. IEEE, 1067–1074.
- [14] Mark Girolami and Ben Calderhead. 2011. Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 73, 2 (2011), 123–214.
- [15] XiaoWei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. 2019. Depth-attentional features for single-image rain removal. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*. 8022–8031.
- [16] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. 2022. Hdr-nerf: High dynamic range neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18398–18408.
- [17] Tai-Xiang Jiang, Ting-Zhu Huang, Xi-Le Zhao, Liang-Jian Deng, and Yao Wang. 2018. Fastderain: A novel video rain streak removal method using directional gradient priors. *IEEE Transactions on Image Processing* 28, 4 (2018), 2089–2102.
- [18] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. 2011. Automatic single-image-based rain streaks removal via image decomposition. *IEEE transactions on image processing* 21, 4 (2011), 1742–1755.
- [19] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* 42, 4 (2023), 1–14.
- [20] Jin-Hwan Kim, Chul Lee, Jae-Young Sim, and Chang-Su Kim. 2013. Single-image deraining using an adaptive nonlocal means filter. In *IEEE international conference on image processing*. IEEE, 914–917.
- [21] Diederik Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations (ICLR)*. San Diego, CA, USA.
- [22] Minghan Li, Qi Xie, Qian Zhao, Wei Wei, Shuhang Gu, Jing Tao, and Deyu Meng. 2018. Video rain streak removal by multiscale convolutional sparse coding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6644–6653.
- [23] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. 2018. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European conference on computer vision (ECCV)*. 254–269.
- [24] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. 2016. Rain streak removal using layer priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2736–2744.
- [25] Beibei Lin, Yeying Jin, Wending Yan, Wei Ye, Yuan Yuan, Shunli Zhang, and Robby T Tan. 2024. NightRain: Nighttime Video Deraining via Adaptive-Rain-Removal and Adaptive-Correction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 3378–3385.
- [26] Yu Luo, Yong Xu, and Hui Ji. 2015. Removing rain from a single image via discriminative sparse coding. In *Proceedings of the IEEE international conference on computer vision*. 3397–3405.
- [27] Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V Sander. 2022. Deblur-nerf: Neural radiance fields from blurry images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12861–12870.
- [28] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. 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.
- [29] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing scenes as neural radiance fields for view synthesis. In *The European Conference on Computer Vision (ECCV)*.
- [30] Siqi Ni, Xueyun Cao, Tao Yue, and Xuemei Hu. 2021. Controlling the rain: From removal to rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6328–6337.
- [31] Naama Pearl, Tali Treibitz, and Simon Korman. 2022. Nan: Noise-aware nerfs for burst-denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12672–12681.
- [32] Rui Qian, Robby T Tan, Wenhan Yang, Jiajuan Su, and Jiaying Liu. 2018. Attentive generative adversarial network for raindrop removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2482–2491.
- [33] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. 2019. On the spectral bias of neural networks. In *International conference on machine learning*. PMLR, 5301–5310.
- [34] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. 2021. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In *Proceedings of the IEEE/CVF international conference on computer vision*. 14335–14345.
- [35] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. 2019. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3937–3946.
- [36] Can Wang, Menglei Chai, Mingming He, Dongdong Chen, and Jing Liao. 2022. Clip-nerf: Text-and-image driven manipulation of neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3835–3844.
- [37] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. 2020. A model-driven deep neural network for single image rain removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3103–3112.
- [38] Hong Wang, Zongsheng Yue, Qi Xie, Qian Zhao, Yefeng Zheng, and Deyu Meng. 2021. From rain generation to rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14791–14801.
- [39] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson WH Lau. 2019. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 12270–12279.
- [40] Bowen Wen, Jonathan Tremblay, Valts Blukis, Stephen Tyre, Thomas Müller, Alex Evans, Dieter Fox, Jan Kautz, and Stan Birchfield. 2023. Bundlesdf: Neural 6-dof tracking and 3d reconstruction of unknown objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 606–617.
- [41] Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. 2022. Image de-raining transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).
- [42] Tao Yan, Mingyue Li, Bin Li, Yang Yang, and Rynson W. H. Lau. 2023. Rain Removal From Light Field Images With 4D Convolution and Multi-Scale Gaussian Process. *IEEE Transactions on Image Processing* 32 (2023), 921–936.
- [43] Wending Yan, Lu Xu, Wenhan Yang, and Robby T Tan. 2022. Feature-aligned video raindrop removal with temporal constraints. *IEEE Transactions on Image Processing* 31 (2022), 3440–3448.
- [44] Wenhan Yang, Jiaying Liu, and Jiashi Feng. 2019. Frame-consistent recurrent video deraining with dual-level flow. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1661–1670.
- [45] Wenhan Yang, Robby T Tan, Shiqi Wang, and Jiaying Liu. 2020. Self-learning video rain streak removal: When cyclic consistency meets temporal correspondence. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 1720–1729.
- [46] Rajeev Yasarla, Vishwanath A Sindagi, and Vishal M Patel. 2020. Syn2real transfer learning for image deraining using gaussian processes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2726–2736.
- [47] Lin Yen-Chen. 2020. NeRF-pytorch. <https://github.com/yenchenlin/nerf-pytorch/>.

929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044

1045	[48] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. 2021. pixelnerf: Neural radiance fields from one or few images. In <i>Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition</i> . 4578–4587.	[51] He Zhang and Vishal M Patel. 2018. Density-aware single image de-raining using a multi-stream dense network. In <i>Proceedings of the IEEE conference on computer vision and pattern recognition</i> . 695–704.	1103
1046			1104
1047	[49] Changfeng Yu, Yi Chang, Yi Li, Xile Zhao, and Luxin Yan. 2021. Unsupervised image deraining: Optimization model driven deep cnn. In <i>Proceedings of the 29th ACM International Conference on Multimedia</i> . 2634–2642.	[52] Kaihao Zhang, Dongxu Li, Wenhan Luo, Wenqi Ren, and Wei Liu. 2023. Enhanced spatio-temporal interaction learning for video deraining: faster and better. <i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i> 45, 1 (2023), 1287–1293.	1105
1048			1106
1049	[50] Zongsheng Yue, Jianwen Xie, Qian Zhao, and Deyu Meng. 2021. Semi-supervised video deraining with dynamical rain generator. In <i>Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition</i> . 642–652.	[53] Mi Zhenxing and Dan Xu. 2022. Switch-nerf: Learning scene decomposition with mixture of experts for large-scale neural radiance fields. In <i>International Conference on Learning Representations (ICLR)</i> .	1107
1050			1108
1051			1109
1052			1110
1053			1111
1054			1112
1055			1113
1056			1114
1057			1115
1058			1116
1059			1117
1060			1118
1061			1119
1062			1120
1063			1121
1064			1122
1065			1123
1066			1124
1067			1125
1068			1126
1069			1127
1070			1128
1071			1129
1072			1130
1073			1131
1074			1132
1075			1133
1076			1134
1077			1135
1078			1136
1079			1137
1080			1138
1081			1139
1082			1140
1083			1141
1084			1142
1085			1143
1086			1144
1087			1145
1088			1146
1089			1147
1090			1148
1091			1149
1092			1150
1093			1151
1094			1152
1095			1153
1096			1154
1097			1155
1098			1156
1099			1157
1100			1158
1101			1159
1102			1160