

DerainNeRF: 3D Scene Estimation with Adhesive Waterdrop Removal

Yunhao Li^{1,2}, Jing Wu², Lingzhe Zhao² and Peidong Liu^{2*}

Abstract— When capturing images through the glass during rainy or snowy weather conditions, the resulting images often contain waterdrops adhered on the glass surface, and these waterdrops significantly degrade the image quality and performance of many computer vision algorithms. To tackle these limitations, we propose a method to reconstruct the clear 3D scene implicitly from multi-view images degraded by waterdrops. Our method exploits an attention network to predict the location of waterdrops and then train a Neural Radiance Fields to recover the 3D scene implicitly. By leveraging the strong scene representation capabilities of NeRF, our method can render high-quality novel-view images with waterdrops removed. Extensive experimental results on both synthetic and real datasets show that our method is able to generate clear 3D scenes and outperforms existing state-of-the-art (SOTA) image adhesive waterdrop removal methods.

I. INTRODUCTION

3D scene representation and estimation techniques have wide range of applications in autonomous driving, robotics, virtual reality (VR) and cultural heritage preservation. In recent years, Neural Radiance Fields (NeRF) [1] have gained popularity for 3D reconstruction and scene representation due to their ability to provide continuous scene representation, robustness in handling complex scenes, and state-of-the-art performance on novel-view image synthesis. However, in many real-world scenarios, particularly those involving outdoor images like autonomous driving and drones, it is likely that the images taken under rainy or snowy weather conditions come with adhesive raindrops, as illustrated in the top row of Fig. 1. Those images will further degrade the performance of related applications, such as 3D reconstruction [1], visual perception [2] [3], object detection [4] [5], and tracking [6] [7]. Therefore, to tackle those limitations, we propose to simultaneously remove the adhesive waterdrops from captured images and recover the underlying 3D scene implicitly in this work, by leveraging the impressive 3D scene representation capability of NeRF.

The vanilla NeRF framework is not robust to images with adhesive waterdrops. Although many researchers have proposed dedicated methods to make NeRF robust under in-the-wild scenarios (e.g. challenging illumination conditions [8],

This work was supported in part by NSFC under Grant 62202389, in part by a grant from the Westlake University-Muyuan Joint Research Institute, and in part by the Westlake Education Foundation.

*Corresponding author

¹Yunhao Li is with School of Computer Science and Engineering, Zhejiang University, No. 866 Yuhangtang Road, Hangzhou, Zhejiang, China. yunhaoli@zju.edu.cn

²Yunhao Li, Jing Wu, Lingzhe Zhao and Peidong Liu are with School of Engineering, Westlake University, No. 600 Dunyu Road, Hangzhou, Zhejiang, China. {wujing05, zhaolingzhe, liupeidong}@westlake.edu.cn

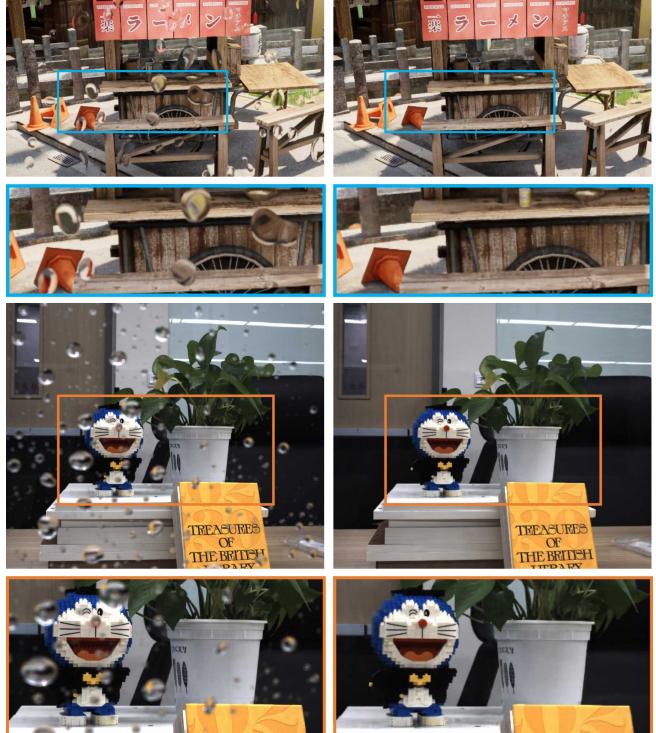


Fig. 1: Given a set of waterdrop images (left column), our DerainNeRF estimates 3D scenes and removes the adhesive waterdrops altogether. It synthesizes clear images (right column) with high quality.

dynamic environments [9]), they cannot handle the images with adhesive waterdrops well because of the random spatial distribution, irregular shapes, and complicated refraction and reflection properties of waterdrops. To address the issue of waterdrop removal, we propose a NeRF-based framework that simultaneously estimates the 3D scenes while removing waterdrops. We refer this model as DerainNeRF, i.e., NeRF with waterdrop removal function.

Inspired by existing waterdrop removal methods, and dedicated NeRF frameworks that handle image occlusion [10] and scene object removal [11] [12], DerainNeRF combines a waterdrop detection network and NeRF for 3D scene representation learning. In particular, it first exploits a pre-trained deep waterdrop detector to predict the locations of waterdrops. It then excludes the waterdrop covered pixels during the training of NeRF, so that it can recover the clear scenes from non-occluded pixels. We evaluate DerainNeRF using both synthetic and real datasets. The experimental results demonstrate that DerainNeRF effectively estimates clear 3D scenes from waterdrop images, and renders novel-

view clear images as is shown in Fig. 1. Both the quantitative and qualitative results demonstrate our method delivers superior quality compared to the existing state-of-the-art image waterdrop removal methods. To the best of our knowledge, DerainNeRF is the first NeRF-based method which takes waterdrop degraded images as input and recovers the clear scene implicitly.

II. RELATED WORK

We review two main areas of the prior works: Neural Radiance Fields (NeRF) and image adhesive waterdrop removal, which are the most related to our work.

A. NeRF

Mildenhall et al. [1] propose NeRF, an epochal 3D scene representation and novel-view image rendering method. Different from discrete voxel-based methods [13], NeRF estimates a continuous scene volumetric function via a Multi Layer Perceptron (MLP). With the help of differentiable volumetric rendering technique, NeRF demonstrates impressive novel view image synthesis performance and 3D scene representation capability. To make NeRF robust to more complicated real-world environments, researchers have proposed many dedicated NeRF variants recently. Some researches extend NeRF to make it compatible for large scale scene reconstruction and representation, making it available to autonomous driving and robotics related applications [14] [15] [16]. Others focus on developing NeRF which can deal with inaccurate camera poses and challenging imaging conditions [17] [18] [19] [20] [21]. There are also many other NeRF variants for High Dynamic Range (HDR) image modeling [22] and NeRF for scene editing [23] [11] [12].

Pan et al. [24] develop a specialized NeRF-based framework which deals with scene of bounded volume and boundary from multi-view posed images with refractive object silhouettes by extending sampling techniques to drawing samples along a curved path modeled by Eikonal equation [25]. WaterNeRF [26] introduced by Sethuraman et al. deals with underwater scene reconstruction by estimating parameters of a physics-based model for image formation. Wang et al. [27] propose NeReF, which estimates the refractive fluid surface with implicit representation.

For occlusion removal and scene editing, Weder et al. [11] propose a NeRF-based scene object removal scheme. Given a user-generated mask, the proposed method first blocks the removed object in input images, impaints the blocked regions using a pre-trained network, then a confidence-driven view-selection scheme enforces multi-view consistencies from the inpainted images. The closest work to ours is Zhu et al. [10], which propose a NeRF-based method with occlusion removal. However, different from focusing on scenarios where occluders are fixed to the scene and camera takes multi-view images, our work can deal with not only the case where the adhesive waterdrops are fixed to the scenes, but also the case when the waterdrops are fixed to the camera. Additionally, our work utilizes a pre-trained rain-detection mechanism from state-of-the-art image waterdrop removal

methods, which is significantly simpler than that of Zhu et al. [10], which requires additional scene MLP and depth-based mask MLP to separate clear background.

B. Image Adhesive Waterdrop removal

Many existing deraining methods focus on removing rain streaks, which has simpler image formation models. However, the physical properties of adhesive waterdrops differ from rain streaks. Earliest methods focus on modeling the waterdrops by estimating its geometric shape, refraction and reflection properties [28] [29]. Others leverages temporal features [30], for example, optical flow, or spatial features [31] [32], for example, disparities, to separate the waterdrops from the clear background. In recent years, some waterdrop removal methods have been introduced. Eigen et al. [33] proposes the first end-to-end image adhesive waterdrop removal method based on deep CNN. Due to its relatively simple and shallow network architecture, its performance becomes poor when the area of adhesive waterdrops in the images becomes large. Qian et al. [34] introduce AttGAN, a waterdrop removal model based on generative adversarial network (GAN) [35]. In AttGAN, the generator contains an attention module, which generates an attention map, then removes the waterdrops based on both input image and attention map. However, the attention-based mechanism in AttGAN fails to leverage global spatial information. Quan et al. [36] develop an approach with shape-driven and channel attention modules, which performs better on large waterdrop removal, but it is still limited to local attention aggregation. Wen et al. [37] propose a video/multi-image waterdrop removal methods for complex driving scenes based on spatial-temporal information fusion by self-attention mechanism and a cross-modality training strategy.

III. METHOD

In this paper, we propose a method to remove waterdrops from multi-view images. When taking images with waterdrop adhered on the glass, we observe two typical scenarios: the waterdrops remain fixed in the scene while the camera is moving, and the waterdrops are static relative to the camera. The former case is often encountered when users take images through a window covered by waterdrops, while the latter is common for the cameras installed on the autonomous vehicles where the waterdrops may fall on the lens. Our method is able to effectively handle both cases. Fig. 2 presents the overview of our method. It first detects the image regions covered by waterdrops from a deep waterdrop detector, and then excludes those regions from the training of NeRF. We will detail each component as follows.

A. Background on NeRF

Given a set of input multi-view images (together with both the camera intrinsic and extrinsic parameters), NeRF [1] first transfers the pixels in the input images into rays using the estimated camera poses from structure from motion (SfM) [38] [39]. It then samples points along each ray, and takes 5D vectors (i.e., the 3D position of sampled point and the

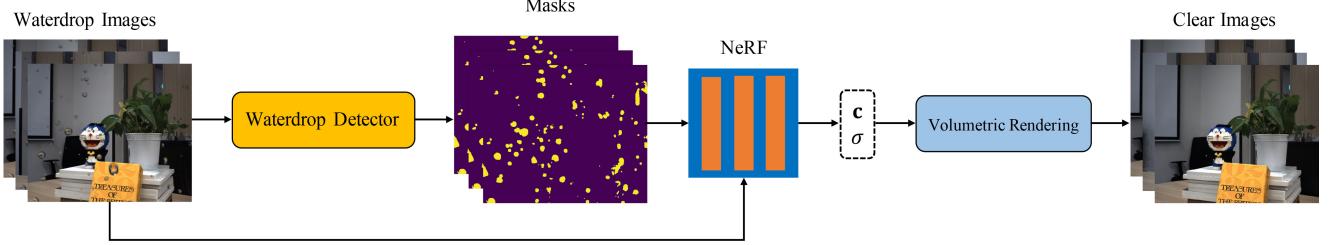


Fig. 2: Training procedure of DerainNeRF. A pre-trained deep waterdrop detector detects waterdrops in input images and generate binary masks, then DerainNeRF utilizes the masks to block waterdrop regions in input images during NeRF training.

2D viewing directions) as input. The volume density σ and view-dependent RGB color c are then estimated by a Multi-layer Perceptron (MLP). The reason that NeRF predicts color from both position and viewing direction is to better deal with the specular reflection of the scene. After obtaining the volume density and color of each sampled point along the ray, it employs a conventional volumetric rendering technique to integrate the density and color to synthesize the corresponding pixel intensity \hat{C} of the image. The whole process can be formally defined via following equation:

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

where t_n and t_f are near and far bounds in volumetric rendering respectively, $\mathbf{r}(t)$ is the sampled 3D point along the ray \mathbf{r} at the distance t from the camera center, $\sigma(\mathbf{r}(t))$ represents the predicted density of the sampled point $\mathbf{r}(t)$ by the MLP, $T(t)$ denotes the accumulated transmittance along the ray from t_n to t , and is defined as $\exp(-\int_{t_n}^t \sigma(\mathbf{r}(s))ds)$, \mathbf{d} is the viewing direction in the world coordinate frame, and $\mathbf{c}(\mathbf{r}(t), \mathbf{d})$ is the predicted color of the sampled point $\mathbf{r}(t)$ by the MLP.

The photometric loss, i.e. the mean squared error (MSE) between the rendered pixel intensity and the real captured intensity, is usually used to train the networks:

$$\mathcal{L} = \sum_{\mathbf{r} \in \mathcal{R}} \left\| \hat{C}(\mathbf{r}) - C(\mathbf{r}) \right\|^2, \quad (2)$$

where both $C(\mathbf{r})$ and $\hat{C}(\mathbf{r})$ denote the real captured and rendered pixel intensities for ray \mathbf{r} respectively, and \mathcal{R} denotes the set of sampled rays.

B. AttGAN

For raindrop detector, we employ the AttGAN model proposed by Qian et al. [34]. AttGAN is a GAN-based single image waterdrop removal network. The generator of AttGAN incorporates a waterdrop detection module based on long-short term memory network (LSTM) [40], and a waterdrop removal module based on U-net [41]. While AttGAN achieves satisfactory performance in waterdrop detection, its waterdrop removal module suffers from certain deficiencies, resulting in suboptimal output images. For example, AttGAN cannot fully eliminate the distortions caused by large waterdrops, leaving the watermark-like effect. In our work, we adopt a pre-trained waterdrop detection module from the AttGAN model as our waterdrop detector.

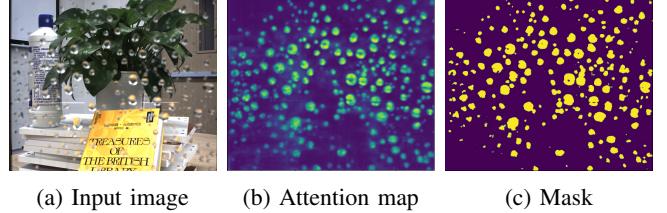


Fig. 3: An example of (a) waterdrop image, (b) attention map and (c) generated binary mask from attention map

C. 3D Scene Estimation from Waterdrop Images

The proposed DerainNeRF takes multi-view waterdrop images as input, then recovers the scene without effect of waterdrops. To achieve this, we feed the input waterdrop images into the pre-trained waterdrop detector, i.e. AttGAN as mentioned in previous section. The detector returns an attention map, where the attention value $\mathbf{A}(u, v) \in [0, 1]$ of pixel at (u, v) indicates the probability whether it is covered by the waterdrop. We then train a NeRF to estimate scenes and block the waterdrop-covered pixels using the binary masks generated from attention maps, as shown in Fig. 2. For each image, we generate a binary mask \mathbf{M} from the following equation

$$\mathbf{M}(u, v) = f(\mathbf{A}(u, v), t) \quad (3)$$

where f is a binary thresholding function, $\mathbf{A}(u, v)$ indicates the attention of the pixel locating at (u, v) coordinate, t is a pre-defined threshold. To further improve the quality of the generated masks, we perform an image dilation operation on the generated binary masks. Fig. 3 shows an example result.

With the generated binary masks, we train NeRF by masking the waterdrop-covered pixels in the photometric loss:

$$\mathcal{L} = \sum_{\mathbf{r} \in \mathcal{R}} \left\| (\hat{C}(\mathbf{r}) - C(\mathbf{r})) \odot (\mathbf{1} - \mathbf{M}) \right\|^2 \quad (4)$$

where \odot represents element-wise multiplication, $\mathbf{1}$ is an all-one vector with the same dimension as binary mask \mathbf{M} . For pixels covered by waterdrops, mask value would be 1, and corresponding loss would be 0, which indicates that these pixels will not contribute to the NeRF optimization.

Furthermore, when the waterdrops are adhered on camera lens, the majority of the droplets remain stationary or move

slowly within camera’s field of view. Therefore we can further enhance the generated mask by averaging over multiple consecutive attention maps. This is due to our observation that certain input images might contain areas that a small portion of waterdrops cannot be detected by the network due to the overexposure or underexposure on these pixels. But in other images, such as images taken from another view, the detection returns to normal as the previously overexposed or underexposed pixels are no longer affected. Therefore, when estimating the scenes with waterdrops lying on the camera lens, we calculate an additional attention map. This attention map is the mean of all generated per-frame attention maps, and it serves to compute the binary mask:

$$\mathbf{M}(u, v) = f(\mathbf{A}(u, v), t) \vee f\left(\frac{1}{n} \sum_{i=1}^n \mathbf{A}_i(u, v), t\right) \quad (5)$$

where $\mathbf{A}_i(u, v)$ is the attention map of i -th input image, and \vee denotes logical OR operation. After the training process is complete, we are able to recover the clear implicit 3D scene without waterdrops. Given an arbitrary camera pose, we can render the clear novel view images following (1).

IV. EXPERIMENTS

A. Implementation Details

To obtain the pre-trained waterdrop detector from AttGAN, we first train the AttGAN following the guidelines of Qian et al. [34], from the dataset in AttGAN paper. When generating the binary masks from attention maps, we set the threshold value t between 0.2 and 0.4, depending on the resolution of input images. For DerainNeRF training, we use ADAM optimizer [42] with learning rate decays from 5×10^{-4} to 5×10^{-5} exponentially. We train our model for 200K iterations, with 1024 rays as batch size, on an NVIDIA GeForce RTX 3090 GPU. We use COLMAP [38], a popular SfM/MVS software to estimate camera poses from input images prior to the training procedure.

B. Datasets

To assess the effectiveness of DerainNeRF, we conduct evaluations on both the synthetic and real datasets. The synthetic dataset comes from the Blender scenes used in Deblur-NeRF [20]. We select 5 virtual scenes, then add the physically simulated waterdrops to the scene via Blender and capture the multi-view images from different camera poses. We capture images under two scenarios: (a) scenes with waterdrops fixed to the scene while the camera is moving (denoted as “-move” in the dataset), and (b) scenes with waterdrops fixed to the camera lens.

For the collection of real datasets, we setup a hardware system. We use a HIKROBOT MV-CA050-12UC camera to capture images. During image acquisition process, we place a 3mm thick glass in front of the camera lenses and spray waterdrops on the glass. We simulate both types of scenarios with this setup (i.e., to move camera only or move camera and glass simultaneously). We have also tested the proposed method with outdoor real dataset, where images are taken

from a moving vehicle under rainy weather conditions. In the outdoor dataset, the waterdrop-covered glass is fixed to the camera. Each scene in the synthetic and real dataset comprises 20-25 images.

C. Results

We evaluate the proposed DerainNeRF on both synthetic and real datasets. Since DerainNeRF can render waterdrop-removed clear images from the reconstructed scene thanks to NeRF’s powerful image synthesis capabilities, we compare the performance of DerainNeRF against that of vanilla NeRF, and state-of-the-art (SOTA) image waterdrop removal methods, including AttGAN [34], Quan et al. [36], and Wen et al. [37]. We also compare the performance against that of vanilla NeRF with waterdrop-removed images from prior mentioned methods. We evaluate the performance quantitatively with the commonly used metrics, such as the structural similarity index (SSIM), peak signal to noise ratio (PSNR), and learned perceptual image patch similarity (LPIPS) [43].

The experimental results on the synthetic dataset provides empirical evidence of the efficacy of DerainNeRF in eliminating waterdrops and reconstructing visually clear 3D scenes with high-fidelity images, as shown in both Fig. 4 and Table I. It is noteworthy that, in certain scenes, the structural similarity index (SSIM) of our method’s images does not exceed that of AttGAN. This observation can be attributed to the fact that our approach does not directly generate waterdrop-free images from the input; instead, it utilizes NeRF to render clear images based on the underlying scene representation. While NeRF successfully preserves the majority of scene details, the rendering process may introduce a marginal loss of image information. Nevertheless, our method exhibits significantly superior performance in terms of peak signal-to-noise ratio (PSNR) and learned perceptual image patch similarity (LPIPS).

To evaluate the performance of DerainNeRF on real datasets, we also conduct qualitative comparisons against state-of-the-art methods. Fig. 5 and Fig. 6 illustrate the comparisons between methods, depicting the outcomes for real indoor and outdoor datasets, respectively. Notably, existing state-of-the-art techniques exhibit limitations when confronted with large waterdrops, resulting in noticeable deficiencies in the output images. In contrast, our DerainNeRF surpasses these methods on real datasets by effectively removing waterdrops in various sizes and shapes.

D. Ablation Study

To better analyze the effectiveness of mask enhancement through average attention map, we compare the results with and without mask enhancement procedure described in (5). We conduct comparisons on synthetic *Tanabata*, *Factory* and *Church* dataset, where the waterdrops are static relative to the camera. Fig. 7 shows the qualitative comparisons between (a) DerainNeRF without mask enhancement and (b) full pipeline. Table II presents quantitative comparisons, where the structural similarity index (SSIM), peak signal to noise ratio (PSNR) and learned perceptual image patch similarity

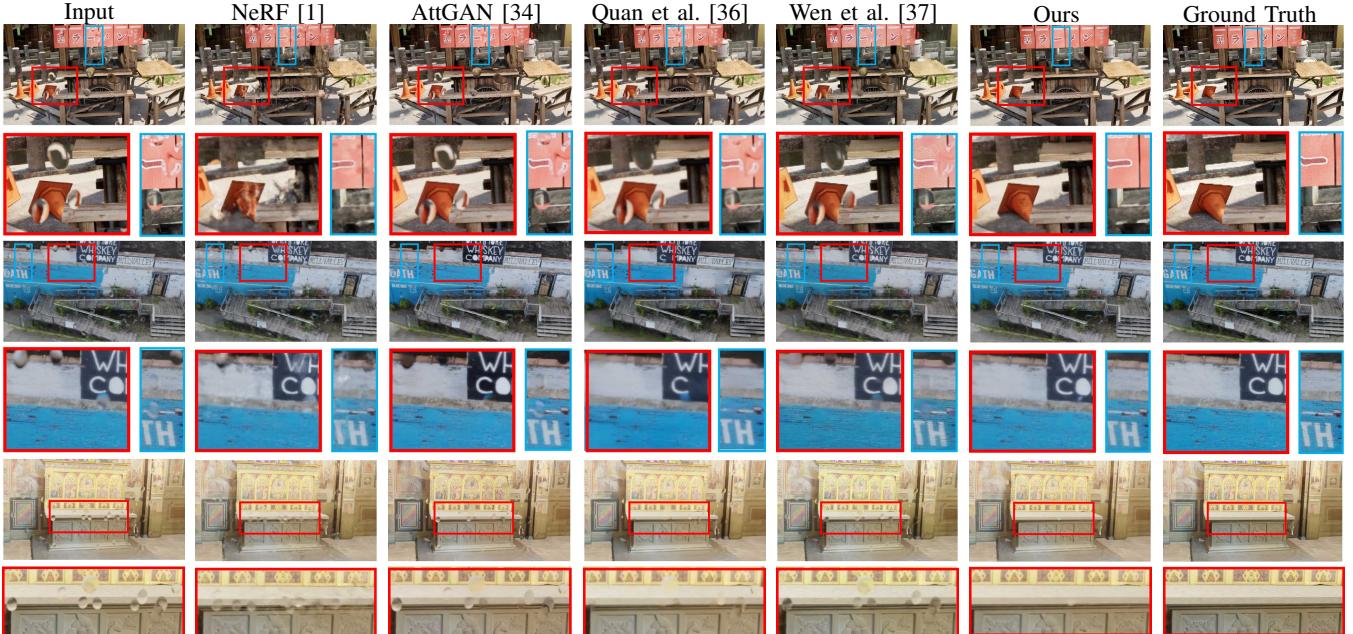


Fig. 4: Qualitative evaluations of our method against SOTA image waterdrop removal methods on the synthetic dataset. Top to bottom shows different scenes including *Tanabata*, *Factory* and *Church*. We render waterdrop-removed images from clear 3D scenes estimated by our method.

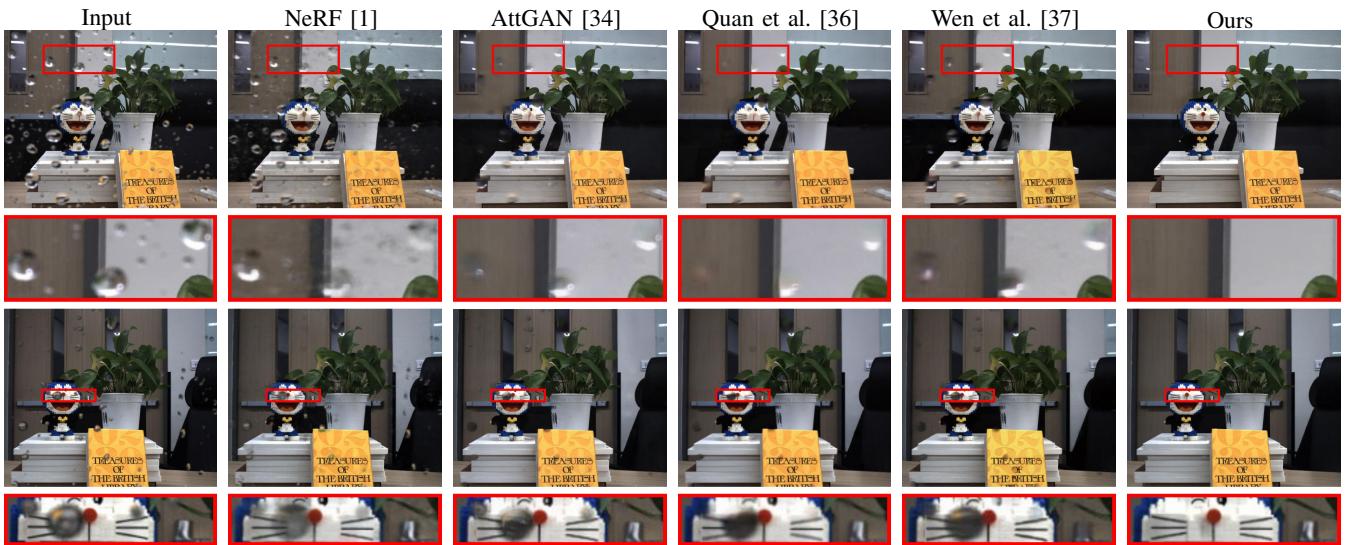


Fig. 5: Qualitative comparisons between different methods with real indoor dataset. The experimental results demonstrate that our method can effectively remove droplets whether the glass with waterdrops is fixed to the scene (top row) or fixed to the camera (bottom row).

TABLE I: Quantitative comparisons on the synthetic dataset. The experimental results demonstrate that our method can render clear images with higher quality than those from existing state-of-the-art image waterdrop removal methods.

	Tanabata			Tanabata-move			Factory			Church			Church-move		
	PSNR↑	SSIM↑	LPIPS↓												
NeRF [1]	19.6145	0.7700	0.2057	23.7986	0.8604	0.1097	26.2478	0.8715	0.1093	27.5391	0.945	0.0921	27.3940	0.9520	0.0817
AttGAN [34]	19.7771	0.8506	0.1359	23.7845	0.9118	0.0775	27.7066	0.9404	0.0854	25.9504	0.9514	0.0752	26.6045	0.9525	0.0760
Quan et al. [36]	20.6615	0.8465	0.1359	23.6408	0.8816	0.1413	25.0735	0.8899	0.1369	27.3451	0.9500	0.0752	26.6356	0.9512	0.1007
Wen et al. [37]	20.9196	0.8399	0.1293	23.2868	0.8784	0.0923	27.5619	0.9153	0.0815	27.7761	0.9462	0.0750	27.6249	0.9536	0.0726
NeRF+AttGAN	20.5404	0.7938	0.1923	23.0334	0.8566	0.0975	26.1938	0.8804	0.0864	26.2998	0.9432	0.0764	27.2996	0.9523	0.0596
NeRF+Quan et al.	20.9633	0.7916	0.2456	23.4550	0.8394	0.1823	26.4362	0.8646	0.1569	27.3061	0.9369	0.1231	27.3506	0.9471	0.1048
NeRF+Wen et al.	21.2252	0.7912	0.1796	23.1937	0.8371	0.1193	27.7594	0.8739	0.1548	28.0618	0.9445	0.0771	27.3503	0.9518	0.0776
DerainNeRF (ours)	26.0367	0.8866	0.1081	24.5683	0.8882	0.0613	30.8504	0.9229	0.0564	30.1028	0.9676	0.0426	28.1776	0.9570	0.0562

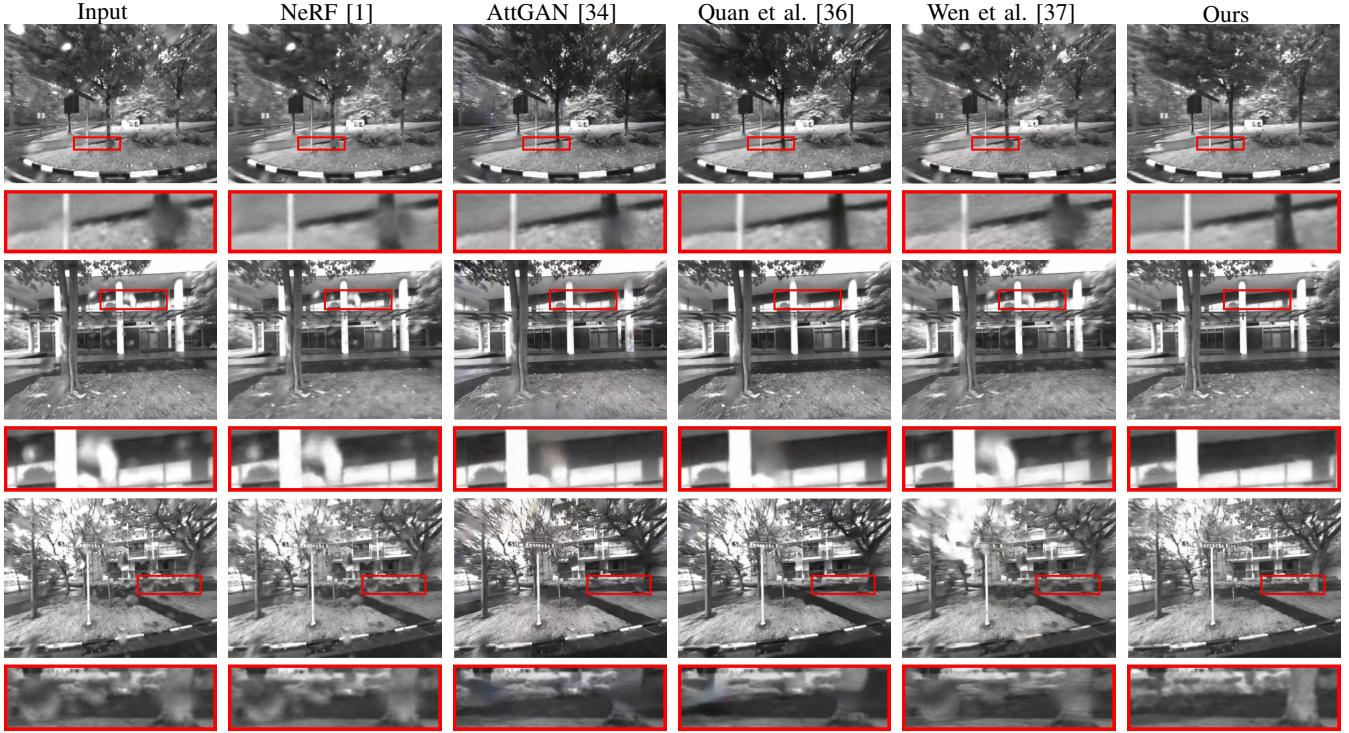


Fig. 6: Qualitative comparisons between different methods on outdoor real dataset. The experimental results demonstrate that our method still presents a satisfying waterdrop removal performance on real outdoor dataset.

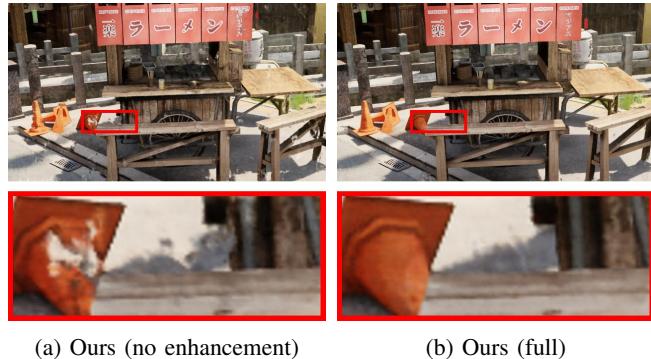


Fig. 7: Qualitative comparison on synthetic dataset for our ablation study.

TABLE II: Ablation study on synthetic *Tanabata*, *Factory* and *Church* dataset.

	Tanabata, Factory and Church		
	PSNR↑	SSIM↑	LPIPS↓
Ours (no enhancement)	27.2611	0.9131	0.0922
Ours (full)	28.3549	0.9299	0.0687

(LPIPS) are the mean value of those on three synthetic datasets. As is shown in Fig. 7 and Table II, the quality of synthesized images from our full pipeline is significantly better than the images rendered from the scene represented by DerainNeRF without mask enhancement. Therefore, mask enhancement based on the average attention map has positive effect during the training of DerainNeRF, especially when the waterdrops become dense.

V. CONCLUSIONS

In this paper, we introduce DerainNeRF, a novel approach for 3D scene estimation from multi-view waterdrop degraded images with NeRF representation. DerainNeRF addresses the challenge of waterdrop removal by utilizing a comprehensive pipeline. Initially, a pre-trained waterdrop detector is employed to identify and localize waterdrops within the input images. Subsequently, our approach estimates clear scenes by leveraging a NeRF-based network exploiting non-occluded pixels. To validate the effectiveness of our proposed method, we conduct a thorough evaluation against existing state-of-the-art techniques for image waterdrop removal with both synthetic and real datasets. The experimental results demonstrate the superior performance of our method in comparison to existing approaches.

REFERENCES

- [1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [2] T. Bruls, W. Maddern, A. A. Morye, and P. Newman, “Mark yourself: Road marking segmentation via weakly-supervised annotations from multimodal data,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1863–1870.
- [3] A. Valada, J. Vertens, A. Dhall, and W. Burgard, “Adapnet: Adaptive semantic segmentation in adverse environmental conditions,” in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 4644–4651.
- [4] Y. Wang and J. Ye, “An overview of 3d object detection,” *arXiv preprint arXiv:2010.15614*, 2020.
- [5] J. Mao, S. Shi, X. Wang, and H. Li, “3d object detection for autonomous driving: A review and new outlooks,” *arXiv preprint arXiv:2206.09474*, 2022.

- [6] R. Qian, X. Lai, and X. Li, “3d object detection for autonomous driving: A survey,” *Pattern Recognition*, vol. 130, p. 108796, 2022.
- [7] E. Arnold, O. Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby, and A. Mouzakitis, “A survey on 3d object detection methods for autonomous driving applications,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3782–3795, 2019.
- [8] R. Martin-Brualla, N. Radwan, M. S. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. Duckworth, “Nerf in the wild: Neural radiance fields for unconstrained photo collections,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7210–7219.
- [9] K. Park, U. Sinha, P. Hedman, J. T. Barron, S. Bouaziz, D. B. Goldman, R. Martin-Brualla, and S. M. Seitz, “Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields,” *arXiv preprint arXiv:2106.13228*, 2021.
- [10] C. Zhu, R. Wan, Y. Tang, and B. Shi, “Occlusion-free scene recovery via neural radiance fields,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 20722–20731.
- [11] S. Weder, G. Garcia-Hernando, A. Monszpart, M. Pollefeys, G. J. Brostow, M. Firman, and S. Vicente, “Removing objects from neural radiance fields,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 16528–16538.
- [12] F. Wei, T. Funkhouser, and S. Rusinkiewicz, “Clutter detection and removal in 3d scenes with view-consistent inpainting,” 2023.
- [13] S. Lombardi, T. Simon, J. Saragih, G. Schwartz, A. Lehrmann, and Y. Sheikh, “Neural volumes: Learning dynamic renderable volumes from images,” *arXiv preprint arXiv:1906.07751*, 2019.
- [14] M. Tancik, V. Casser, X. Yan, S. Pradhan, B. Mildenhall, P. P. Srinivasan, J. T. Barron, and H. Kretzschmar, “Block-nerf: Scalable large scene neural view synthesis,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 8248–8258.
- [15] Y. Xiangli, L. Xu, X. Pan, N. Zhao, A. Rao, C. Theobalt, B. Dai, and D. Lin, “Bungeenerf: Progressive neural radiance field for extreme multi-scale scene rendering,” 2023.
- [16] H. Turki, D. Ramanan, and M. 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*, 2022, pp. 12922–12931.
- [17] Y. Jeong, S. Ahn, C. Choy, A. Anandkumar, M. Cho, and J. Park, “Self-calibrating neural radiance fields,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5846–5854.
- [18] C.-H. Lin, W.-C. Ma, A. Torralba, and S. Lucey, “Barf: Bundle-adjusting neural radiance fields,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 5741–5751.
- [19] Z. Wang, S. Wu, W. Xie, M. Chen, and V. A. Prisacariu, “Nerf+: Neural radiance fields without known camera parameters,” *arXiv preprint arXiv:2102.07064*, 2021.
- [20] L. Ma, X. Li, J. Liao, Q. Zhang, X. Wang, J. Wang, and P. V. Sander, “Deblur-nerf: Neural radiance fields from blurry images,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12861–12870.
- [21] P. Wang, L. Zhao, R. Ma, and P. Liu, “Bad-nerf: Bundle adjusted deblur neural radiance fields,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4170–4179.
- [22] X. Huang, Q. Zhang, Y. Feng, H. Li, X. Wang, and Q. Wang, “Hdr-nerf: High dynamic range neural radiance fields,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18398–18408.
- [23] Y.-J. Yuan, Y.-T. Sun, Y.-K. Lai, Y. Ma, R. Jia, and L. Gao, “Nerf-editing: geometry editing of neural radiance fields,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 18353–18364.
- [24] J.-I. Pan, J.-W. Su, K.-W. Hsiao, T.-Y. Yen, and H.-K. Chu, “Sampling neural radiance fields for refractive objects,” in *SIGGRAPH Asia 2022 Technical Communications*, 2022, pp. 1–4.
- [25] A. R. Bruss, “The eikonal equation: Some results applicable to computer vision,” *Journal of Mathematical Physics*, vol. 23, no. 5, pp. 890–896, 1982.
- [26] A. V. Sethuraman, M. S. Ramanagopal, and K. A. Skinner, “Water-nerf: Neural radiance fields for underwater scenes,” *arXiv preprint arXiv:2209.13091*, 2022.
- [27] Z. Wang, W. Yang, J. Cao, L. Xu, J. Yu, and J. Yu, “Nerf: Neural refractive field for fluid surface reconstruction and implicit representation,” *arXiv preprint arXiv:2203.04130*, 2022.
- [28] K. Garg and S. K. Nayar, “Photometric model of a rain drop,” 2003.
- [29] Y.-J. Yu, H.-Y. Jung, and H.-G. Cho, “A new water droplet model using metaball in the gravitational field,” *Computers & Graphics*, vol. 23, no. 2, pp. 213–222, 1999.
- [30] S. You, R. T. Tan, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi, “Adherent raindrop modeling, detection and removal in video,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 9, pp. 1721–1733, 2015.
- [31] A. Yamashita, I. Fukuchi, and T. Kaneko, “Noises removal from image sequences acquired with moving camera by estimating camera motion from spatio-temporal information,” in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 3794–3801.
- [32] Y. Tanaka, A. Yamashita, T. Kaneko, and K. T. Miura, “Removal of adherent waterdrops from images acquired with a stereo camera system,” *IEICE TRANSACTIONS on Information and Systems*, vol. 89, no. 7, pp. 2021–2027, 2006.
- [33] D. Eigen, D. Krishnan, and R. Fergus, “Restoring an image taken through a window covered with dirt or rain,” in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 633–640.
- [34] R. Qian, R. T. Tan, W. Yang, J. Su, and J. Liu, “Attentive generative adversarial network for raindrop removal from a single image,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2482–2491.
- [35] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in neural information processing systems*, vol. 27, 2014.
- [36] Y. Quan, S. Deng, Y. Chen, and H. Ji, “Deep learning for seeing through window with raindrops,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2463–2471.
- [37] Q. Wen, Y. Wu, and Q. Chen, “Video waterdrop removal via spatio-temporal fusion in driving scenes,” 2023.
- [38] J. L. Schonberger and J.-M. Frahm, “Structure-from-motion revisited,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4104–4113.
- [39] V. Sitzmann, M. Zollhöfer, and G. Wetzstein, “Scene representation networks: Continuous 3d-structure-aware neural scene representations,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [40] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, “Convolutional lstm network: A machine learning approach for precipitation nowcasting,” *Advances in neural information processing systems*, vol. 28, 2015.
- [41] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.
- [42] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2017.
- [43] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.