# NeRFLiX: High-Quality Neural View Synthesis by Learning a Degradation-Driven Inter-viewpoint MiXer

Kun Zhou[1,2*] Wenbo Li[3*] Yi Wang[4]
Tao Hu[3] Nianjuan Jiang[2] Xiaoguang Han[1] Jiangbo Lu[2†]
[1]SSE, CUHK-Shenzhen, [2]SmartMore Corporation [3]CUHK [4]Shanghai AI Laboratory
kunzhou@link.cuhk.edu.cn, {wenboli,taohu}@cse.cuhk.edu.hk
hanxiaoguang@cuhk.edu.cn,{jiangbo.lu,wygamle}@gmail.com


Figure 1. We propose NeRFLiX, a general NeRF-agnostic restorer that is capable of improving neural view synthesis quality. The first example is from Tanks and Temples [25], the second/third examples are from LLFF [36], and the last one is a user scene captured by a mobile phone. RegNeRF-V3 [39] means the model trained with three input views.

## Abstract

*Neural radiance fields (NeRF) show great success in novel view synthesis. However, in real-world scenes, recovering high-quality details from the source images is still challenging for the existing NeRF-based approaches, due to the potential imperfect calibration information and scene representation inaccuracy. Even with high-quality training frames, the synthetic novel views produced by NeRF models still suffer from notable rendering artifacts, such as noise, blur, etc. Towards to improve the synthesis quality of NeRF-based approaches, we propose NeRFLiX, a general NeRF-agnostic restorer paradigm by learning a degradation-driven inter-viewpoint mixer. Specially, we design a NeRF-style degradation modeling approach and construct large-scale training data, enabling the possibility of effectively removing NeRF-native rendering artifacts for existing deep neural networks. Moreover, beyond the degradation removal, we propose an inter-viewpoint aggregation framework that is able to fuse highly related high-quality training images, pushing the performance of cutting-edge NeRF models to entirely new levels and producing highly photo-realistic synthetic views.*

## 1. Introduction

Neural radiance fields (NeRF) can generate photo-realistic images from new viewpoints, playing a heated role in novel view synthesis. In light of NeRF's [37] success, numerous approaches [2, 9, 11, 19, 34, 35, 38, 40, 41, 47, 53, 54, 59, 61, 65] along these lines have been proposed, contin-

*Equal contribution
†Corresponding author

ually raising the performance to greater levels. In fact, one prerequisite of NeRF is the precise camera settings of the taken photos for training [22, 32, 61]. However, accurately calibrating camera poses is exceedingly difficult in practice. Contrarily, the shape-radiance co-adaption issue [74] reveals that while the learned radiance fields can perfectly explain the training views with inaccurate geometry, they poorly generalize to unseen views. On the other hand, the capacity to represent sophisticated geometry, lighting, object materials, and other factors is constrained by the simplified scene representation of NeRF [19, 77, 78]. On the basis of such restrictions, advanced NeRF models may nonetheless result in *notable artifacts* (such as blur, noise, detail missing, and more), which we refer to as *NeRF-style degradations* in this article and are shown in Fig. 1.

To address the aforementioned limitations, numerous works have been proposed. For example, some studies, including [22, 59, 66, 70], jointly optimize camera parameters and neural radiance fields to refine camera poses as precisely as possible in order to address the camera calibration issue. Another line of works [19, 73, 77, 78] presents physical-aware models that simultaneously take into account the object materials and environment lighting, as opposed to using MLPs or neural voxels to implicitly encode both the geometry and appearance. To meet the demands for high-quality neural view synthesis, one has to carefully examine all of the elements when building complex inverse rendering systems. In addition to being challenging to optimize, they are also not scalable for rapid deployment with hard re-configurations in new environments. Regardless of the intricate physical-aware rendering models, *is it possible to design a practical NeRF-agnostic restorer to directly enhance synthesized views from NeRFs*?

In the low-level vision, it is critical to construct large-scale paired data to train a deep restorer for eliminating real-world artifacts [56, 72]. When it comes to NeRF-style degradations, there are two challenges: (1) sizable paired training data; (2) NeRF degradation analysis. First, it is unpractical to gather *large-scale* training pairs (more specifically, raw outputs from well-trained NeRFs and corresponding ground truths). Second, the modeling of NeRF-style degradation has received little attention. Unlike real-world images that generally suffer from JPEG compression, sensor noise, and motion blur, the NeRF-style artifacts are complex and differ from the existing ones. As far as we know, **no** previous studies have ever investigated NeRF-style degradation removal which effectively leverages the ample research on image and video restoration.

In this work, we are motivated to have the *first* study on the feasibility of simulating large-scale NeRF-style paired data, opening the possibility of training a NeRF-agnostic restorer for improving the NeRF rendering frames. To this end, we present a novel degradation simulator for typical NeRF-style artifacts (e.g., rendering noise and blur) considering the NeRF mechanism. We review the overall NeRF rendering pipeline and discuss the typical NeRF-style degradation cases. Accordingly, we present three basic degradation types to simulate the real rendered artifacts of NeRF synthetic views and empirically evaluate the distribution similarity between real rendered photos and our simulated ones. The feasibility of developing NeRF-agnostic restoration models has been made possible by constructing a sizable dataset that covers a variety of NeRF-style degradations, over different scenes.

Next, we show the necessity of our simulated dataset and demonstrate that existing state-of-the-art image restoration frameworks can be used to eliminate NeRF visual artifacts. Furthermore, we notice, in a typical NeRF setup, neighboring high-quality views come for free, and they serve as potential reference bases for video-based restoration with a multi-frame aggregation and fusion module. However, this is not straightforward because NeRF input views are taken from a variety of very different angles and locations, making the estimation of correspondence quite challenging. To tackle this problem, we propose a degradation-driven interviewpoint "mixer" that progressively aligns image contents at the pixel and patch levels. In order to maximize efficiency and improve performance, we also propose a fast view selection technique to only choose the most pertinent reference training views for aggregation, as opposed to using the entire NeRF input views.

In a nutshell, we present a NeRF-agnostic restorer (termed NeRFLiX) which learns a degradation-driven interviewpoint mixer. As illustrated in Fig. 1, given NeRF synthetic frames with various rendering degradations, NeRFLiX successfully restores high-quality results. Our contributions are summarized as

- **Universal enhancer for NeRF models.** NeRFLiX is powerful and adaptable, removing NeRF artifacts and restoring clearly details, pushing the performance of cutting-edge NeRF models to entirely new levels.
- **NeRF rendering degradation simulator.** We develop a NeRF-style degradation simulator (NDS), constructing massive amounts of paired data and aiding the training of deep neural networks to improve the quality of NeRF-rendered images.
- **Inter-viewpoint mixer.** Based on our constructed NDS, we further propose an inter-viewpoint baseline that is able to *mix* high-quality neighboring views for more effective restorations.
- **Training time acceleration.** We show how NeRFLiX makes it possible for NeRF models to produce even *better* results with a *50*% reduction in training time.

## 2. Related Works

**NeRF-based novel view synthesis.** NeRF-based novel view synthesis has received a lot of attention recently and has been thoroughly investigated. For the first time, Mildenhall *et al*. [37] propose the neural radiance field to implicitly represent static 3D scenes and synthesize novel views from multiple posed images. Inspired by their successes, a lot of NeRF-based models [2, 8, 11, 13, 19–21, 23, 26, 33–35, 38, 41, 42, 44, 47, 51, 54, 62, 65, 73, 76] have been proposed. For example, point-NeRF [63] and DS-NeRF [14] incorporate sparse 3D point cloud and depth information for eliminating the geometry ambiguity for NeRFs, achieving more accurate/efficient 3D point sampling and better rendering quality. Plenoxels [16], TensoRF [7], DirectVoxGo [45], FastNeRF [17], Plenoctrees [67], KiloNeRF [43], and Mobilenerf [10], aim to use various advanced technologies to speed up the training or inference phases. Though these methods have achieved great progress, due to the potential issue of inaccurate camera poses, simplified pinhole camera models as well as scene representation inaccuracy, they still suffer from rendering artifacts of the predicted novel views.

**Degradation simulation.** Since no existing works have explored the NeRF-style degradation cases, we will overview the real-world image restoration works that are most related to ours. The previous image/video super-resolution approaches [15,27,28,31,55,57,68,79,79,80] typically follow a fix image degradation type (e.g., blur, bicubic/bilinear down-sampling). Due to the large domain shift between the real-world and simulated degradations, the earlier image restoration methods [27,29,71,79] generally fail to remove complex artifacts of the real-world images. In contrast, BSRGAN [72] design a practical degradation approach for real-world image super-resolution. In their degradation process, multiple degradations are considered and applied in random orders, largely covering the diversity of real-world degradations. Compared with the previous works, BSRGAN achieves much better results quantitatively and qualitatively. Real-ESRGAN [56] develops a second-order degradation process for real-world image super-resolution. In this work, we propose a NeRF-style degradation simulator and construct a large-scale training dataset for modeling the NeRF rendering artifacts.

**Correspondence estimation.** In the existing literature, the video restoration methods [3, 6, 48, 52, 69] aim to restore a high-quality frame from multiple low-quality frames. To achieve this goal, cross-frame correspondence estimation is essential to effectively aggregate the informative temporal contents. Some works [5, 6, 64, 69] explore building pixel-level correspondences through optical-flow estimation and perform frame-warping for multi-frame compensation. Another line of works [49, 55, 80] tries to use deformable convolution networks (DCNs [12]) for adaptive correspondence estimation and aggregation. More recently, transformer-based video restoration models [4, 30] implement spatial-temporal aggregation through an attention mechanism and achieve promising performance. However, it is still challenging to perform accurate correspondence estimation between frames captured with very distinctive viewpoints.

## 3. Preliminaries

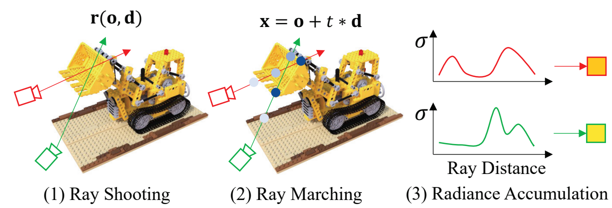


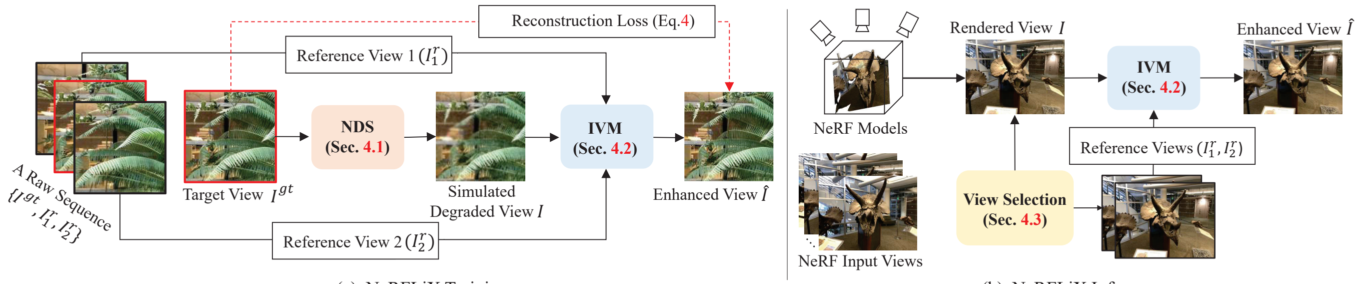

Figure 2. A general illustration of NeRF-based novel view synthesis pipeline. Three main steps are involved: (1) ray shooting, (2) ray marching, and (3) radiance accumulation.

In this section, we review the general pipeline of NeRF-based novel view synthesis and discuss potential rendering artifacts. As shown in Fig. 2, three main steps are involved in the rendering: (1) Ray Shooting. To render the color of a target pixel in a particular view, NeRF utilizes the camera's calibrated parameters $\pi$ to generate a ray $\mathbf{r}(\mathbf{o}, \mathbf{d})$ through this pixel, where $\mathbf{o}, \mathbf{d}$ are the camera center and the ray direction. (2) Ray Marching. A set of 3D points are sampled along the chosen ray as it moves across the 3D scene represented by neural radiance fields. The NeRF models encode a 3D scene and predict the colors and densities of these points. (3) Radiance Accumulation. The pixel color is extracted by integrating the predicted radiance features of the sampled 3D points.

**Discussion.** We can see that establishing a relationship between 2D photos and the 3D scene requires camera calibration. Unfortunately, it is very challenging to precisely calibrate the camera poses, leading to noisy 3D sampling. Meanwhile, some previous works [22, 59, 66, 70] also raise other concerns, including the non-linear pinhole camera model [22] and shape-radiance ambiguity [74]. Because of these inherent limitations, as discussed in Section 1, NeRF models still synthesize unsatisfied novel test views.

## 4. Methodology

**Overview.** In this work, we present NeRFLiX, a general NeRF-agnostic restorer which employs a degradation-driven inter-viewpoint mixer to enhance novel view images rendered by NeRF models. It is made up of two essential components: a NeRF-style degradation simulator (NDS) and an inter-viewpoint mixer (IVM). As seen in Fig. 3(a), during the training phase, we employ the proposed NDS




(a) NeRFLiX Training (b) NeRFLiX Inference

Figure 3. Illustration of our proposed NeRFLiX. It consists of two essential modules: (1) NeRF degradation simulator that constructs paired training data $\{I, I_1^r, I_2^r | I^{gt}\}$ from a raw sequence $\{I^{gt}, I_1^r, I_2^r\}$, (2) inter-viewpoint mixer trained on this simulated data is capable of restoring high-quality frames from NeRF rendered views.

to create large-scale paired training data, which are subsequently used to train an IVM for improving a NeRF-rendered view using two corresponding reference pictures (reference views). In the inference stage, as illustrated in Fig. 3(b), IVM is adopted to enhance a rendered view by fusing useful information from the selected most relevant reference views.

### 4.1. NeRF-Style Degradation Simulator (NDS)

Due to the difficulties in gathering well-posed scenes under various environments and training NeRF models for each scene, it is infeasible to directly collect large amounts of *paired* NeRF data for artifact removal. To address this challenge, motivated by BSRGAN [72], we design a general NeRF degradation simulator to produce a sizable training dataset that is visually and statistically comparable to NeRF-rendered images (views).

To begin with, we collect raw data from LLFF-T[1] and Vimeo90K [64] where the adjacent frames are treated as raw sequences. Each raw sequence consists of three images $\{I^{gt}, I_1^r, I_2^r\}$: a target view $I^{gt}$ and its two reference views $\{I_1^r, I_2^r\}$. To construct the paired data from a raw sequence, we use the proposed NDS to degrade $I^{gt}$ and obtain a simulated degraded view $I$, as shown in Fig. 3(a).

The degradation pipeline is illustrated in Fig 4. We design three types of degradation for a target view $I^{gt}$: splatted Gaussian noise (SGN), re-positioning (Re-Pos.), and anisotropic blur (A-Blur). It should be noted that there *may be other models for such a simulation*, and we only utilize this route to evaluate and justify the feasibility of our idea.

**Splatted Gaussian noise.** Although additive Gaussian noise is frequently employed in image/video denoising, NeRF rendering noise clearly differs. Rays that hit a 3D point will be re-projected within a nearby 2D area because of noisy camera parameters. As a result, the NeRF-style noise is dispersed over a 2D space. This observation led us

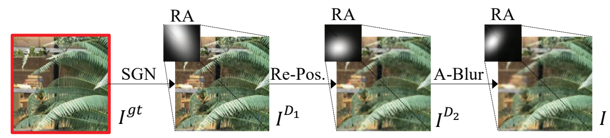



Figure 4. Overview of our NDS pipeline: using our proposed degradations, we process a target view $I^{gt}$ to produce its simulated degraded view $I$. "SGN", "Re-Pos." and "A-Blur" refer to the splatted Gaussian, re-positioning, anisotropic blur degradations, and "RA" is the region adaptive strategy.

to present a splatted Gaussian noise, which is defined as

$$I^{D1} = (I^{gt} + n) \circledast g, \tag{1}$$

where $n$ is a 2D Gaussian noise map with the same resolution as $I^{gt}$ and $g$ is an isotropic Gaussian blur kernel.

**Re-positioning.** We design a re-positioning degradation to simulate ray jittering. We add a random 2D offset $\delta_i, \delta_j \in [-2, 2]$ with probability 0.1 for a pixel at location $(i, j)$

$$I^{D2}(i,j) = \begin{cases} I^{D1}(i,j) & \text{if } \ p > 0.1 \\ I^{D1}(i+\delta_i, j+\delta_j) & \text{else } \ p \leq 0.1 \end{cases} \tag{2}$$

where $p$ is uniformly distributed in $[0, 1]$.

**Anisotropic blur.** Additionally, from our observation, NeRF synthetic frames also contain blurry contents. To simulate blur patterns, we use anisotropic Gaussian kernels to blur the target frame.

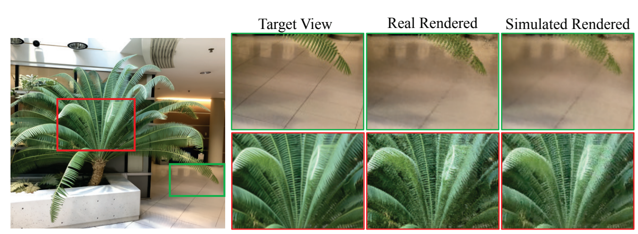



Figure 5. A visual example of real and simulated rendered views.

**Region adaptive strategy.** Neural radiance fields are often supervised with unbalanced training views. As a result,

[1] the training parts of LLFF [36].

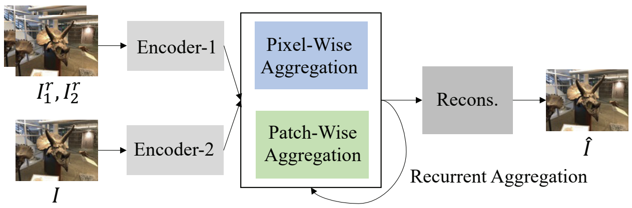



Figure 6. The framework of our inter-viewpoint mixer.

given a novel view, the projected 2D areas have varying degradation levels. Thus, we carry out each of the employed degradations in a spatially variant manner. More specifically, we define a mask $M$ as a two-dimensional oriented anisotropic Gaussian [18]

$$M(i,j) = G(i - c_i, j - c_j; \sigma_i, \sigma_j, A), \tag{3}$$

where $(c_i, c_j), (\sigma_i, \sigma_j)$ are the means and standard deviations and $A$ is an orientation angle. After that, we use the mask $M$ to linearly blend the input and output of each degradation, finally achieving region-adaptive degradations. As shown in Fig. 5, our simulated rendered views visually match the real NeRF-rendered ones. All the detailed settings of NDS are elaborated in our supplementary materials.

At last, with our NDS, we can obtain a great number of training pairs, and each paired data consists of two high-quality reference views $\{I_1^r, I_2^r\}$, a simulated degraded view $I$, and the corresponding target view $I^{gt}$. Next, we show how the constructed paired data $\{I, I_1^r, I_2^r | I^{gt}\}$ can be used to train our IVM.

### 4.2. Inter-viewpoint Mixer (IVM)

**Problem formulation.** Given a degraded view $I$ produced by our NDS or NeRF models, we aim to extract useful information from its two high-quality reference views $\{I_1^r, I_2^r\}$ and restore an enhanced version $\hat{I}$.

**IVM architecture.** For multi-frame processing, existing techniques either use optical flow [5, 52, 69] or deformable convolutions [12, 30, 55] to realize the correspondence estimation and aggregation for *consistent* displacements. In contrast, NeRF rendered and input views come from very different angles and locations, making it challenging to perform precise inter-viewpoint aggregation.

To address this problem, we propose IVM, a hybrid recurrent inter-viewpoint "mixer" that progressively fuses pixel-wise and patch-wise contents from two high-quality reference views, achieving more effective inter-viewpoint aggregation. There are three modules i.e., feature extraction, hybrid inter-viewpoint aggregation and reconstruction, as shown in Fig. 6. Two convolutional encoders are used in the feature extraction stage to process the degraded view $I$ and two high-quality reference views $\{I_1^r, I_2^r\}$, respectively. We then use inter-viewpoint window-based attention modules and deformable convolutions to achieve recurrent patch-wise and pixel-wise aggregation. Finally, the enhanced view $\hat{I}$ is generated using the reconstruction module under the supervision

$$Loss = |\hat{I} - I^{gt}|, \text{where } \hat{I} = f(I, I_1^r, I_2^r; \theta), \tag{4}$$

where $\theta$ is the learnable parameters of IVM. The framework architecture is given in our supplementary materials.

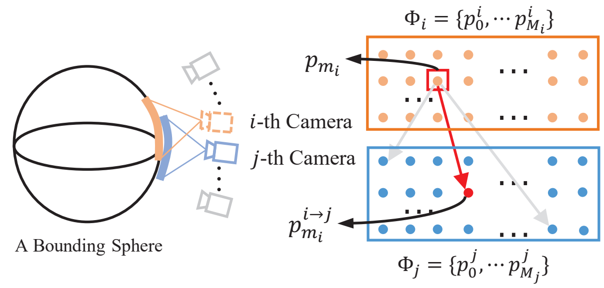



Figure 7. Illustration of our view selection strategy.

### 4.3. View Selection

In the inference stage, for a NeRF-rendered view $I$, our IVM produces an enhanced version by aggregating contents from two neighboring high-quality views. But, multiple input views are available and only a part of them are largely overlapped with $I$. In general, only the most pertinent input views are useful for the inter-viewpoint aggregation.

To this end, we develop a view selection strategy to choose two reference views $\{I_1^r, I_2^r\}$ from the input views that are most overlapped with the rendered view $I$. Specifically, we formulate the view selection problem based on the pinhole camera model. An arbitrary 3D scene can be roughly approximated as a bounding sphere in Fig. 7, and cameras are placed around it to take pictures. When camera-emitted rays hit the sphere, there are a set of intersections. We refer to the 3D point sets as $\Phi_i = \{p_0^i, p_1^i, \cdots, p_{M_i}^i\}$ and $\Phi_j = \{p_0^j, p_1^j, \cdots, p_{M_j}^j\}$ for the $i$-th and $j$-th cameras. For $m_i$-th intersection $p_{m_i}^i \in \Phi_i$ of view $i$, we search its nearest point in view $j$ with the L2 distance

$$p_{m_i}^{i \to j} = \underset{p \in \Phi_j}{\arg\min}(||p - p_{m_i}^i||_2^2). \tag{5}$$

Then the matching cost from the $i$-th view to the $j$-th view is calculated by

$$C_{i \to j} = \sum_{m_i=0}^{M_i} ||p_{m_i}^i - p_{m_i}^{i \to j}||_2^2. \tag{6}$$

We finally obtain the mutual matching cost between views $i$ and $j$ as

$$C_{i \leftrightarrow j} = C_{i \to j} + C_{j \to i}. \tag{7}$$

In this regard, two reference views $\{I_1^r, I_2^r\}$ are selected at the least mutual matching costs for enhancing the NeRF-rendered view $I$. Note that we also adopt this strategy to decide the two reference views for the LLFF-T [36] data during the training phase.

| Method | PSNR (dB)↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|
| TensoRF [7] (ECCV'22) | 26.73 | 0.839 | 0.204 |
| TensoRF [7] + NeRFLiX | **27.39** (↑ 0.66) | **0.867** | **0.149** |
| Plenoxels [16] (CVPR'22) | 26.29 | 0.839 | 0.210 |
| Plenoxels [16] + NeRFLiX | **26.90** (↑ 0.61) | **0.864** | **0.156** |
| NeRF-mm [59] (ARXIV'21) | 22.98 | 0.655 | 0.440 |
| NeRF-mm [59] + NeRFLiX | **23.38** (↑ 0.40) | **0.694** | **0.360** |
| NeRF [37] (ECCV'20) | 26.50 | 0.811 | 0.250 |
| NeRF [37] + NeRFLiX | **27.26**(↑ 0.76) | **0.863** | **0.159** |

(a) Quantitative results on the LLFF under LLFF-P1.

| Method | PSNR (dB)↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|
| NLF [1] (CVPR'22) | 27.46 | 0.868 | 0.136 |
| NLF [1] + NeRFLiX | **28.19** (↑ 0.73) | **0.899** | **0.093** |
| RegNeRF-V3 [39] (CVPR'22) | 19.10 | 0.587 | 0.373 |
| RegNeRF-V3 [39] + NeRFLiX | **19.68** (↑ 0.58) | **0.661** | **0.260** |
| RegNeRF-V6 [39] (CVPR'22) | 23.06 | 0.759 | 0.242 |
| RegNeRF-V6 [39] + NeRFLiX | **23.90**(↑ 0.84) | **0.815** | **0.144** |
| RegNeRF-V9 [39] (CVPR'22) | 24.81 | 0.818 | 0.196 |
| RegNeRF-V9 [39] + NeRFLiX | **25.68** (↑ 0.87) | **0.863** | **0.114** |

(b) Quantitative results on the LLFF [36] under LLFF-P2. RegNeRF-V3(6,9) takes 3(6,9) input views for training.

Table 1. Quantitative analysis of our NeRFLiX on LLFF [36].

## 5. Experiments

### 5.1. Implementation Details

We train the IVM for 300K iterations. The batch size is 16 and the patch size is 128. We adopt random cropping, vertical or horizontal flipping, and rotation augmentations. Apart from the inherent viewpoint changes over $\{I, I_1^r, I_2^r\}$, random offsets ($\pm 5$ pixels) are globally applied to the two reference views ($I_1^r, I_2^r$) to model more complex motion. We adopt an Adam [24] optimizer and a Cosine annealing strategy to decay the learning rate from $5 \times 10^{-4}$ to 0. We train a single IVM on the LLFF-T and Vimeo datasets and test it on all benchmarks (including user-captured scenes).

### 5.2. Datasets and Metrics

We conduct the experiments on three widely used datasets, including LLFF [36], Tanks and Temples [25], and Noisy LLFF Synthetic.

**LLFF [36].** LLFF is a real-world dataset, where 8 different scenes have 20 to 62 images. Following the commonly used protocols [1, 7, 16, 39, 59], we adopt $1008 \times 756$ resolution for *LLFF-P1* and $504 \times 376$ resolution for *LLFF-P2*.

**Tanks and Temples [25].** It contains 5 scenes captured by inward-facing cameras. There are 152-384 images in the $1920 \times 1080$ resolution. It should be noted that the viewpoints of different frames are significantly larger than LLFF.

**Noisy LLFF Synthetic [37].** There are 8 virtual scenes, each of which has 400 images with a size of $800 \times 800$. To simulate noisy in-the-wild calibration, we ad hoc apply camera jittering (random rotation and translation are employed) to the precise camera poses.

**Metrics.** Following previous NeRF methods, we adopt PSNR (↑)/SSIM [58] (↑)/LPIPS [75](↓) for evaluation.

### 5.3. Improvement over SOTA NeRF Models

We demonstrate the effectiveness of our approach by showing that it consistently improves the performance of cutting-edge NeRF approaches across various datasets.

**LLFF.** In order to fully verify the generalization ability of our NeRFLiX, we investigate six representative models, including NeRF [37], TensoRF [7], Plenoxels [16], NeRF-mm [59], NLF [1], and RegNeRF [39]. Using rendered images of NeRF models as inputs to our model, we aim to further improve the synthesis quality. The quantitative results are provided in Table 1. We find that under both of the two protocols, our method raises NeRF model performance entirely to new levels. For example, NeRFLiX improves Plenoxels [16] by *0.61dB/0.025/0.054* in terms of PSNR/SSIM/LPIPS.

| Method | PSNR (dB)↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|
| TensoRF [7] (ECCV'22) | 28.43 | 0.920 | 0.142 |
| TensoRF [7] + NeRFLiX | **28.94**(↑ 0.51) | **0.930** | **0.120** |
| DIVeR [60] (CVPR'22) | 28.16 | 0.913 | 0.145 |
| DIVeR [60] + NeRFLiX | **28.61**(↑ 0.45) | **0.924** | **0.127** |

(a) Improvement over TensoRF and DIVeR on Tanks and Temples.

| Method | PSNR (dB)↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|
| TensoRF [7] (ECCV'22) | 22.83 | 0.881 | 0.147 |
| TensoRF [7] + NeRFLiX | **24.12** (↑ 1.29) | **0.913** | **0.092** |
| Plenoxels [16] (CVPR'22) | 23.69 | 0.882 | 0.127 |
| Plenoxels [16] + NeRFLiX | **25.51** (↑ 1.82) | **0.920** | **0.084** |

(b) Improvement over TensoRF and Plenoxels on noisy LLFF Synthetic.

| Method | PSNR (dB)↑/SSIM↑/LPIPS↓ |
|---|---|
| TensoRF [7](4 hours) | 26.73/ 0.839/ 0.204 |
| TensoRF [7](**2 hours**) | 26.18/ 0.819/ 0.230 |
| [7](**2 hours**) + NeRFLiX | **27.14/ 0.858/ 0.165** |
| Plenoxels [16](24 minutes) | 26.29/ 0.839/ 0.210 |
| Plenoxels [16](**10 minutes**) | 25.73/ 0.804/ 0.252 |
| [16](**10 minutes**) + NeRFLiX | **26.60/ 0.847/ 0.181** |

(c) Improvement over TensoRF and Plenoxels trained with half of the recommended iterations on LLFF [36] under LLFF-P1.

Table 2. Quantitative evaluation of the improvement of NeRFLiX for various NeRFs.

**Tanks and Temples.** Due to large variations of camera viewpoints, even advanced NeRF models, e.g., TensoRF [7] and DIVeR [60], show obviously inferior rendering quality on this dataset. As illustrated in Table 2a, we show that our NeRFLiX can still boost the performance of these models by a large margin, especially TensoRF [7] achieves *0.51dB/0.01/0.022* improvements on PSNR/SSIM/LPIPS.

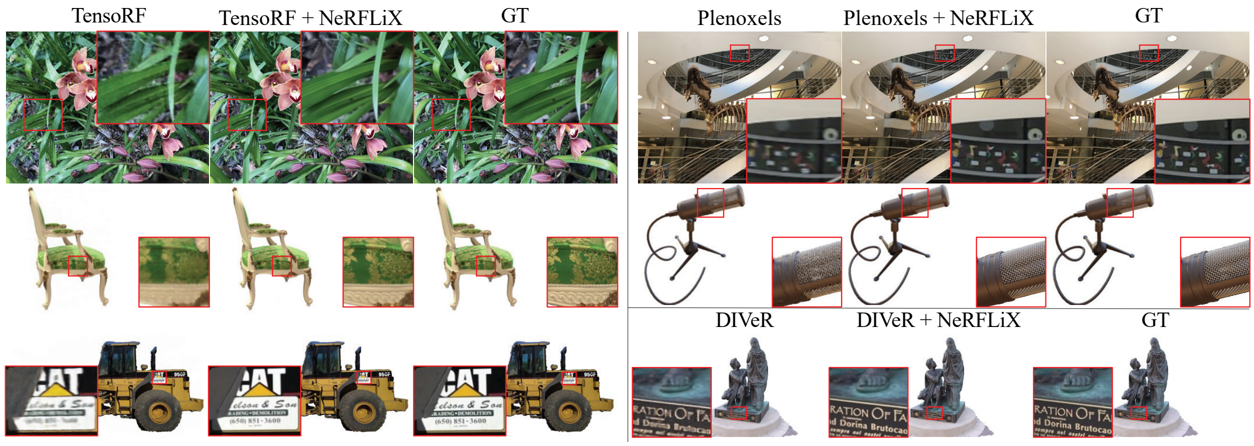



Figure 8. Qualitative evaluation of the improvement over three SOTA NeRFs on LLFF, Noisy LLFF Synthetic, and Tanks and Temples.

**Noisy LLFF Synthetic.** Apart from in-the-wild benchmarks above, we also demonstrate the enhancement capability of our model on noisy LLFF Synthetic. From the results shown in Table 2b, we see that our NeRFLiX yields substantial improvements for two SOTA NeRF models.

**Qualitative results.** In Fig. 8, we provide some visual examples for qualitative assessment. It is obvious that our NeRFLiX restores clearer image details while removing the majority of NeRF-style artifacts in the rendered images, clearly manifesting the effectiveness of our method. More results are provided in our supplementary materials.

## 5.4. Training Acceleration for NeRF Models

In this section, we show how our approach makes it possible for NeRF models to produce better results even with a 50% reduction in training time. To be more precise, we use NeRFLiX to improve the rendered images of two SOTA NeRF models after training them with half the training period specified in the publications. The enhanced results *outperform* the counterparts with full-time training, as shown in Table 2c. Notably, NeRFLiX has reduced the training period for Plenoxels [16] from 24 minutes to 10 minutes while also improving the quality of the rendered images.

## 5.5. Ablation Study

In this section, we conduct comprehensive experiments on LLFF [36] under the LLFF-P1 protocol to analyze each of our designs. We use TensoRF [7] as our baseline[2].

### 5.5.1 NeRF-Style Degradation Simulator

**Simulation quality.** We first examine the simulation quality of the proposed NeRF-style degradation simulator. To this end, we analyze the distribution of our degraded images, BSR [72] degraded images and NeRF rendered images on LLFF [36]. We use t-SNE [50] to visualize deep

[2]The TensoRF results (26.70dB/0.838/0.204) we examined differ slightly from the published results (26.73dB/0.839/0.204).

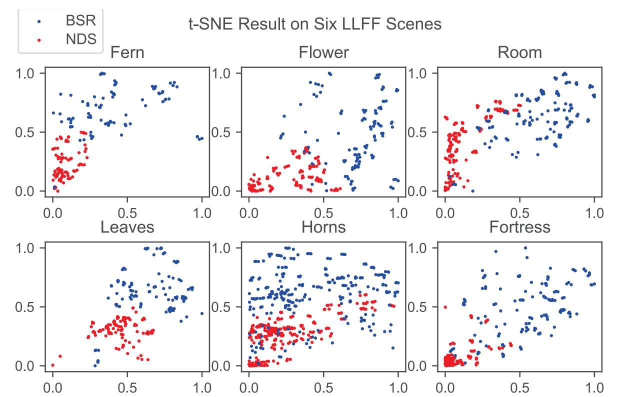



Figure 9. Quantitative comparison between our NDS and BSR [72] over six LLFF scenes. We draw the normalized differences between the simulated images of the two degradation methods and the real NeRF-rendered images. The smaller values, the better results are (best viewed in color).

image features (by Inception-v3 [46]) and results are shown in Fig. 9. Our simulated data is statistically much closer to the real rendered images than BSR. This conclusion is also supported by Table 3b, which demonstrates that our NDS significantly surpasses BSR and yields 0.6-1.0dB improvements when used for learning NeRF degradations.

**Degradation type.** We also evaluate the detailed contribution of each data degradation. We use simulated data to train our models by gradually including four types of degradation, as illustrated in Table 4. From quantitative comparisons on LLFF [36], we observe that all employed degradations are beneficial to our system.

### 5.5.2 Inter-viewpoint Mixer

**View selection strategy.** We develop a view selection strategy to make full use of high-quality reference views. As shown in Fig. 10, our system can identify the most relevant views for quality enhancement when compared to random selection. Also, the quantitative results in Table 5 sug-

| Model | TensoRF(Base) | $\text{SwIR}_\mathbf{N}$ | $\text{DATSR}_\mathbf{N}$ | $\text{EDVR}_\mathbf{N}$ | $\text{VST}_\mathbf{N}$ |
|---|---|---|---|---|---|
| PSNR | 26.70 | 26.82 | 26.84 | 26.88 | 26.79 |
| SSIM | 0.838 | 0.845 | 0.843 | 0.847 | 0.842 |

(a) Quantitative results of the improvements using existing image/video processing models trained on our simulated dataset.

| Models | BSR | NDS | SwIR | IVM | PSNR | SSIM |
|---|---|---|---|---|---|---|
| $\text{SwIR}_\mathbf{B}$ | ✓ | | ✓ | | 26.20 | 0.834 |
| $\text{SwIR}_\mathbf{N}$ | | ✓ | ✓ | | **26.82** | **0.845** |
| $\text{IVM}_\mathbf{B}$ | ✓ | | | ✓ | 26.40 | 0.842 |
| $\text{IVM}_\mathbf{N}$ | | ✓ | | ✓ | **27.39** | **0.867** |

(b) Comparison of our NDS and the BSR degradation models [72].

Table 3. Quantitative comparison between our NDS and BSR [72]. We use subscripts $_\mathbf{N}$, $_\mathbf{B}$, to represent the models trained with our NDS dataset and BSR, respectively.

| Models | SGN | Re-Pos. | A-Blur | RA | PSNR(dB) | SSIM |
|---|---|---|---|---|---|---|
| Model-1 | ✓ | | | | 27.08 | 0.856 |
| Model-2 | ✓ | ✓ | | | 27.13 | 0.858 |
| Model-3 | ✓ | ✓ | ✓ | | 27.21 | 0.859 |
| Model-4 | ✓ | ✓ | ✓ | ✓ | **27.39** | **0.867** |

Table 4. Influences of different degradations used in our NeRF-style degradation simulator. "SGN" and "RA" are shorted for splatted Gaussian noise and region-adaptive schemes and "A-Blur" refers to anisotropic Gaussian blur.

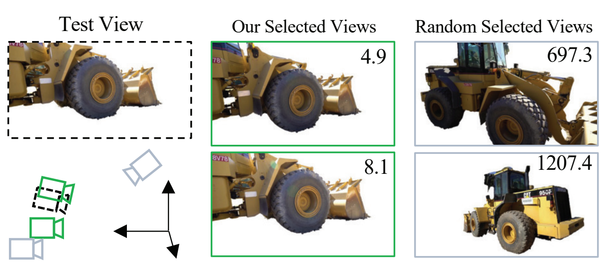



Figure 10. Comparison of the two view selection methods. The number in each sub-figure is the matching cost calculated in Eq. 7.

| Method | LLFF | Tanks and Temple |
|---|---|---|
| Random | 27.06dB/ 0.856 | 28.51dB/ 0.925 |
| View Selection | **27.39dB/ 0.867** | **28.94dB/ 0.930** |

Table 5. Ablation studies of our view selection strategy.

gest that our view selection achieves significantly improved results, illustrating the usefulness of our method.

**Hybrid recurrent multi-view aggregation.** To handle large viewpoint differences in reference and rendered views, we develop a hybrid recurrent inter-viewpoint aggregation network. We train models using either pixel-wise or patch-wise aggregation and test different iterations to assess the proposed IVM. Models using a single aggregation approach, as illustrated in Table 6, perform worse than our full configuration. Additionally, by gradually increasing the iteration numbers from 1 to 3, we achieve improvements of 0.12dB and 0.06dB, albeit at an additional cost of 66 ms and 46 ms for aggregation. Last, compared with the exist-

| Method | PSNR(dB) | SSIM | LPIPS | Speed (ms) |
|---|---|---|---|---|
| Pixel-wise | 27.13 | 0.862 | 0.179 | 230 |
| Patch-wise | 27.00 | 0.854 | 0.183 | 237 |
| Hybrid + R1 | 27.21 | 0.865 | 0.173 | 181 |
| Hybrid + R2 | 27.33 | 0.866 | 0.157 | 247 |
| Hybrid + R3 | **27.39** | **0.867** | **0.149** | 293 |

Table 6. Ablation studies of hybrid inter-viewpoint aggregation module. The running time is tested with an input size of $256\times256$.

ing state-of-the-art models in Table 3a, thanks to the recurrent hybrid aggregation strategy, our IVM outperforms all of these models in terms of quantitative results, demonstrating the strength of our aggregation design.

**Limitation.** Though NeRFLiX achieves promising progress for its universal improvements over existing NeRF models. There are still some future directions that deserve further exploration. (1) Our NDS is one of many possible solutions for NeRF degradation simulation. (2) Exploring real-time inter-viewpoint mixers is interesting and useful.

## 6. Conclusion

We presented NeRFLiX, a general NeRF-agnostic restoration paradigm for high-quality neural view synthesis. We systematically analyzed the NeRF rendering pipeline and introduced the concept of NeRF-style degradations. Towards to eliminate NeRF-style artifacts, we presented a novel NeRF-style degradation simulator and constructed a large-scale simulated dataset. Benefiting from our simulated dataset, we demonstrated how SOTA deep neural networks could be trained for NeRF artifact removal. To further restore missing details of NeRF rendered frames, we proposed an inter-viewpoint mixer that is capable of aggregating multi-view frames captured from free viewpoints. Additionally, we developed a view selection scheme for choosing the most pertinent reference frames, largely alleviating the computing burden while achieving superior results. Extensive experiments have verified the effectiveness of our NeRFLiX. Code will be made publicly available.

**Acknowledgments.** This work is partially supported by Shenzhen Science and Technology Program KQTD20210811090149095, the Basic Research Project No. HZQB-KCZYZ-2021067 of Hetao Shenzhen-HK S&T Cooperation Zone. It was also supported in part by NSFC-62172348 and Shenzhen General Project with No. JCYJ20220530143604010, the National Key R&D Program of China with grant No. 2018YFB1800800, by Shenzhen Outstanding Talents Training Fund 202002, by Guangdong Research Projects No. 2017ZT07X152 and No. 2019CX01X104, by the Guangdong Provincial Key Laboratory of Future Networks of Intelligence (Grant No. 2022B1212010001), and by Shenzhen Key Laboratory of Big Data and Artificial Intelligence (Grant No. ZDSYS201707251409055).

## References


[1] Benjamin Attal, Jia-Bin Huang, Michael Zollhöfer, Johannes Kopf, and Changil Kim. Learning neural light fields with ray-space embedding networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 6

[2] 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. 1, 3

[3] Jiezhang Cao, Yawei Li, Kai Zhang, and Luc Van Gool. Video super-resolution transformer. *arXiv*, 2021. 3

[4] Mingden Cao, Yanbo Fan, Yong Zhang, Jue Wang, and Yujiu Yang. Vdtr: Video deblurring with transformer. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022. 3

[5] Kelvin CK Chan, Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Basicvsr: The search for essential components in video super-resolution and beyond. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 3, 5

[6] Kelvin CK Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. Basicvsr++: Improving video super-resolution with enhanced propagation and alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5972–5981, 2022. 3

[7] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision (ECCV)*, 2022. 3, 6, 7

[8] Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao Su. Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14124–14133, 2021. 3

[9] Tianlong Chen, Peihao Wang, Zhiwen Fan, and Zhangyang Wang. Aug-nerf: Training stronger neural radiance fields with triple-level physically-grounded augmentations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15191–15202, 2022. 1

[10] Zhiqin Chen, Thomas Funkhouser, Peter Hedman, and Andrea Tagliasacchi. Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures. *arXiv preprint arXiv:2208.00277*, 2022. 3

[11] Forrester Cole, Kyle Genova, Avneesh Sud, Daniel Vlasic, and Zhoutong Zhang. Differentiable surface rendering via non-differentiable sampling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6088–6097, 2021. 1, 3

[12] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 764–773, 2017. 3, 5

[13] Chenxi Lola Deng and Enzo Tartaglione. Compressing explicit voxel grid representations: fast nerfs become also small. *arXiv preprint arXiv:2210.12782*, 2022. 3

[14] 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. 3

[15] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 3

[16] 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. 3, 6, 7

[17] Stephan J Garbin, Marek Kowalski, Matthew Johnson, Jamie Shotton, and Julien Valentin. Fastnerf: High-fidelity neural rendering at 200fps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14346–14355, 2021. 3

[18] J-M Geusebroek, Arnold WM Smeulders, and Joost Van De Weijer. Fast anisotropic gauss filtering. *IEEE transactions on image processing*, 12(8):938–943, 2003. 5

[19] Yuan-Chen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. Nerfren: Neural radiance fields with reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18409–18418, 2022. 1, 2, 3

[20] Tao Hu, Shu Liu, Yilun Chen, Tiancheng Shen, and Jiaya Jia. Efficientnerf efficient neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12902–12911, 2022. 3

[21] Jeffrey Ichnowski*, Yahav Avigal*, Justin Kerr, and Ken Goldberg. Dex-NeRF: Using a neural radiance field to grasp transparent objects. In *Conference on Robot Learning (CoRL)*, 2020. 3

[22] Yoonwoo Jeong, Seokjun Ahn, Christopher Choy, Anima Anandkumar, Minsu Cho, and Jaesik Park. Self-calibrating neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5846–5854, 2021. 2, 3

[23] Mohammad Mahdi Johari, Yann Lepoittevin, and François Fleuret. Geonerf: Generalizing nerf with geometry priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18365–18375, 2022. 3

[24] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[25] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)*, 36(4):1–13, 2017. 1, 6

[26] Andreas Kurz, Thomas Neff, Zhaoyang Lv, Michael Zollhöfer, and Markus Steinberger. Adanerf: Adaptive sampling for real-time rendering of neural radiance fields. In

*European Conference on Computer Vision*, pages 254–270. Springer, 2022. 3

[27] Wenbo Li, Kun Zhou, Lu Qi, Nianjuan Jiang, Jiangbo Lu, and Jiaya Jia. Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. *Advances in Neural Information Processing Systems*, 33:20343–20355, 2020. 3

[28] Wenbo Li, Kun Zhou, Lu Qi, Liying Lu, and Jiangbo Lu. Best-buddy gans for highly detailed image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 1412–1420, 2022. 3

[29] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3867–3876, 2019. 3

[30] Jingyun Liang, Jiezhang Cao, Yuchen Fan, Kai Zhang, Rakesh Ranjan, Yawei Li, Radu Timofte, and Luc Van Gool. Vrt: A video restoration transformer. *arXiv preprint arXiv:2201.12288*, 2022. 3, 5

[31] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 3

[32] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. Barf: Bundle-adjusting neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5741–5751, 2021. 2

[33] Haotong Lin, Sida Peng, Zhen Xu, Hujun Bao, and Xiaowei Zhou. Efficient neural radiance fields with learned depth-guided sampling. *arXiv preprint arXiv:2112.01517*, 2021. 3

[34] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel 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*, pages 7210–7219, 2021. 1, 3

[35] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. 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*, pages 16190–16199, 2022. 1, 3

[36] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (TOG)*, 38(4):1–14, 2019. 1, 4, 5, 6, 7

[37] 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. In *ECCV*, 2020. 1, 3, 6

[38] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, July 2022. 1, 3

[39] Michael Niemeyer, Jonathan T. Barron, Ben Mildenhall, Mehdi S. M. Sajjadi, Andreas Geiger, and Noha Radwan. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 6

[40] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5865–5874, 2021. 1

[41] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10318–10327, 2021. 1, 3

[42] Daniel Rebain, Wei Jiang, Soroosh Yazdani, Ke Li, Kwang Moo Yi, and Andrea Tagliasacchi. Derf: Decomposed radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14153–14161, 2021. 3

[43] Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14335–14345, 2021. 3

[44] Mohammed Suhail, Carlos Esteves, Leonid Sigal, and Ameesh Makadia. Light field neural rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8269–8279, 2022. 3

[45] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5459–5469, 2022. 3

[46] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016. 7

[47] 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. 1, 3

[48] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *European conference on computer vision*, pages 402–419. Springer, 2020. 3

[49] Yapeng Tian, Yulun Zhang, Yun Fu, and Chenliang Xu. Tdan: Temporally-deformable alignment network for video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3360–3369, 2020. 3

[50] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008. 7

[51] Chen Wang, Xian Wu, Yuan-Chen Guo, Song-Hai Zhang, Yu-Wing Tai, and Shi-Min Hu. Nerf-sr: High quality neural

radiance fields using supersampling. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6445–6454, 2022. 3

[52] Longguang Wang, Yulan Guo, Zaiping Lin, Xinpu Deng, and Wei An. Learning for video super-resolution through hr optical flow estimation. In *Asian Conference on Computer Vision*, pages 514–529. Springer, 2018. 3, 5

[53] Liao Wang, Jiakai Zhang, Xinhang Liu, Fuqiang Zhao, Yanshun Zhang, Yingliang Zhang, Minye Wu, Jingyi Yu, and Lan Xu. Fourier plenoctrees for dynamic radiance field rendering in real-time. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13524–13534, 2022. 1

[54] Qianqian Wang, Zhicheng Wang, Kyle Genova, Pratul P Srinivasan, Howard Zhou, Jonathan T Barron, Ricardo Martin-Brualla, Noah Snavely, and Thomas Funkhouser. Ibrnet: Learning multi-view image-based rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, 2021. 1, 3

[55] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 3, 5

[56] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1905–1914, 2021. 2, 3

[57] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 3

[58] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6

[59] Zirui Wang, Shangzhe Wu, Weidi Xie, Min Chen, and Victor Adrian Prisacariu. Nerf–: Neural radiance fields without known camera parameters. *arXiv preprint arXiv:2102.07064*, 2021. 1, 2, 3, 6

[60] Liwen Wu, Jae Yong Lee, Anand Bhattad, Yu-Xiong Wang, and David Forsyth. Diver: Real-time and accurate neural radiance fields with deterministic integration for volume rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16200–16209, 2022. 6

[61] Zijin Wu, Xingyi Li, Juewen Peng, Hao Lu, Zhiguo Cao, and Weicai Zhong. Dof-nerf: Depth-of-field meets neural radiance fields. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 1718–1729, 2022. 1, 2

[62] Fanbo Xiang, Zexiang Xu, Milos Hasan, Yannick Hold-Geoffroy, Kalyan Sunkavalli, and Hao Su. Neutex: Neural texture mapping for volumetric neural rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7119–7128, 2021. 3

[63] Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. Point-nerf: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5438–5448, 2022. 3

[64] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Freeman. Video enhancement with task-oriented flow. *IJCV*, 127(8):1106–1125, 2019. 3, 4

[65] Guo-Wei Yang, Wen-Yang Zhou, Hao-Yang Peng, Dun Liang, Tai-Jiang Mu, and Shi-Min Hu. Recursive-nerf: An efficient and dynamically growing nerf. *IEEE Transactions on Visualization and Computer Graphics*, 2022. 1, 3

[66] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems*, 33:2492–2502, 2020. 2, 3

[67] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. Plenoctrees for real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5752–5761, 2021. 3

[68] Ke Yu, Xintao Wang, Chao Dong, Xiaoou Tang, and Chen Change Loy. Path-restore: Learning network path selection for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 3

[69] Songhyun Yu, Bumjun Park, Junwoo Park, and Jechang Jeong. Joint learning of blind video denoising and optical flow estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 500–501, 2020. 3, 5

[70] Jiahui Zhang, Fangneng Zhan, Rongliang Wu, Yingchen Yu, Wenqing Zhang, Bai Song, Xiaoqin Zhang, and Shijian Lu. Vmrf: View matching neural radiance fields. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6579–6587, 2022. 2, 3

[71] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3217–3226, 2020. 3

[72] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4791–4800, 2021. 2, 3, 4, 7, 8

[73] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5453–5462, 2021. 2, 3

[74] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020. 2, 3

[75] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of*

*the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018. 6

[76] Wenyuan Zhang, Ruofan Xing, Yunfan Zeng, Yu-Shen Liu, Kanle Shi, and Zhizhong Han. Fast learning radiance fields by shooting much fewer rays. *arXiv preprint arXiv:2208.06821*, 2022. 3

[77] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (TOG)*, 40(6):1–18, 2021. 2

[78] Yuanqing Zhang, Jiaming Sun, Xingyi He, Huan Fu, Rongfei Jia, and Xiaowei Zhou. Modeling indirect illumination for inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18643–18652, 2022. 2

[79] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018. 3

[80] Kun Zhou, Wenbo Li, Liying Lu, Xiaoguang Han, and Jiangbo Lu. Revisiting temporal alignment for video restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6053–6062, 2022. 3