JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

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

WED-PM-002

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-SZ, [2]SmartMore Corporation, [3]CUHK, [4]Shanghai AI Laboratory

Project Page: *https://redrock303.github.io/nerflix/*

Preview

- **NeRFLiX is a general NeRF-agnostic restorer that is capable of improving view synthesis quality.**

➢ NeRF models typically require (1) accurate camera poses, (2) complex inverse rendering systems, (3) illumination and material calibration, and (4) disentanglement of geometry and appearance to synthesize high-quality novel views.

➢ While being free of solving these challenges, NeRFLiX can directly enhance the NeRF's rendering quality by learning a degradation-driven inter-viewpoint mixer .

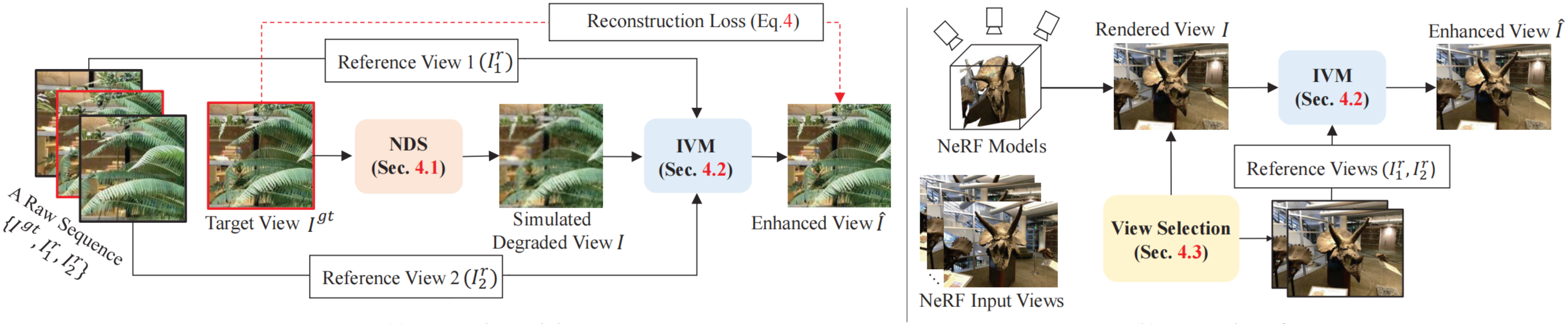
Introduction

Key Ideas:

- We propose a practical NeRF-style degradation simulator (called NDS) to model the NeRF-rendered artifacts and construct large-scale paired training data, enabling the possibility of effectively removing NeRF-native rendering artifacts for deep image/video restorers.
- Taking advantage of the simulated training data, we further develop a hybrid recurrent inter-viewpoint mixer (IVM) to fuse high-quality contents from reference views (the training photos of NeRFs) and eliminate NeRF-rendered degradations/artifacts.

Method

NeRFLiX Pipeline

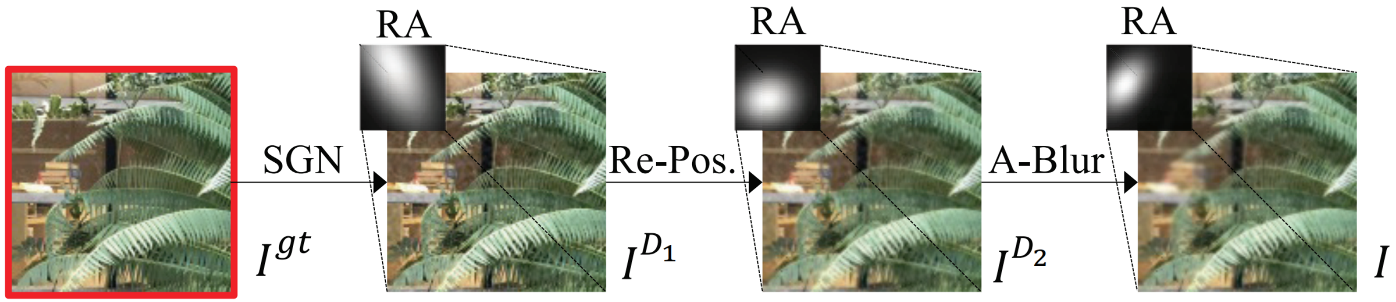


(a) NeRFLiX Training (b) NeRFLiX Inference

- We utilized NDS to generate a large-scale training dataset. By leveraging the large-scale paired samples constructed by NDS, we trained IVM to eliminate NeRF-style artifacts and aggregate noise-free contents from reference views, resulting in a significant enhancement of NeRF-rendered views.

Method

NeRF-style degradation simulator (NDS)



- Splatted Gaussian noise:

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

- Re-positioning:

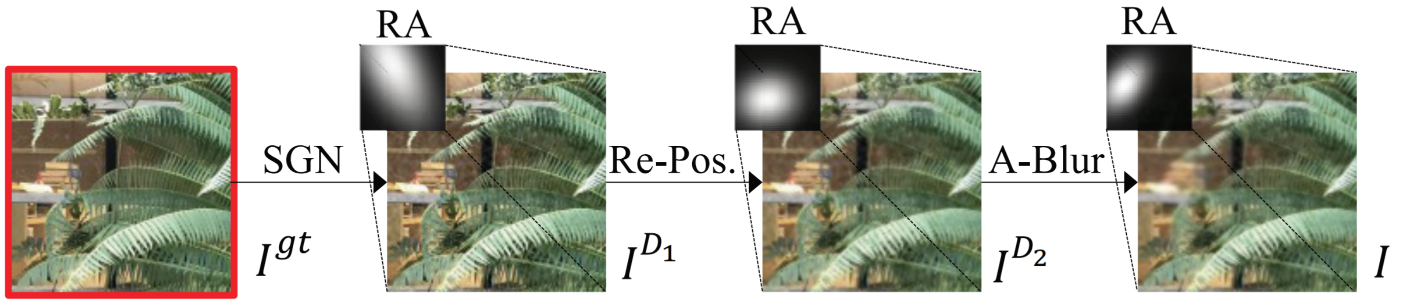
$$I^{D2}(i,j) = \begin{cases} I^{D1}(i,j) & \text{if } \ p > 0.1 \\ I^{D1}(i+\delta_i, j+\delta_j) & \text{else } \ p \leq 0.1 \end{cases}$$

- Anisotropic blur:

$$I^{D3} = I^{D2} \circledast \hat{g}$$

Method

NeRF-style degradation simulator (NDS)



- Region adaptive strategy:

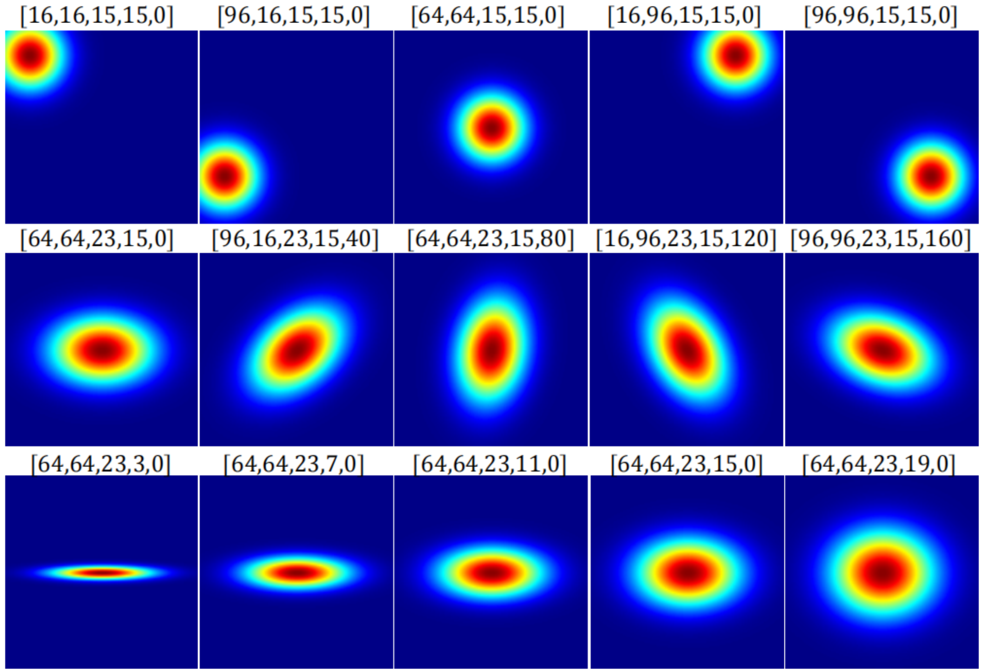
$$M(i,j) = G(i - c_i, j - c_j; \sigma_i, \sigma_j, A)$$

$$I^{D_t} = I^{D_{t-1}} \odot (1 - M_t) + (\mathrm{D_t}(\mathrm{I}^{\mathrm{D}_{t-1}}) \odot M_t)$$

M: Region-adaptive mask

$t \in \{1,2,3\}$ refers to the t-th degradation operation

$\odot$: Element-wise multiplication

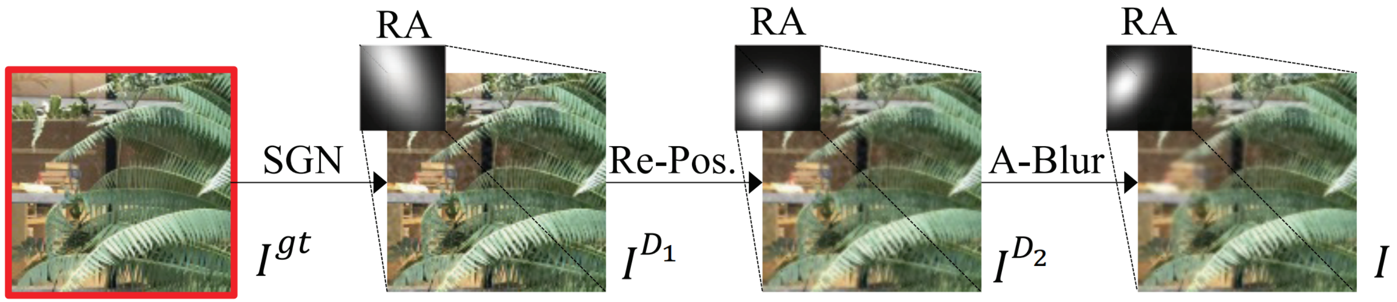


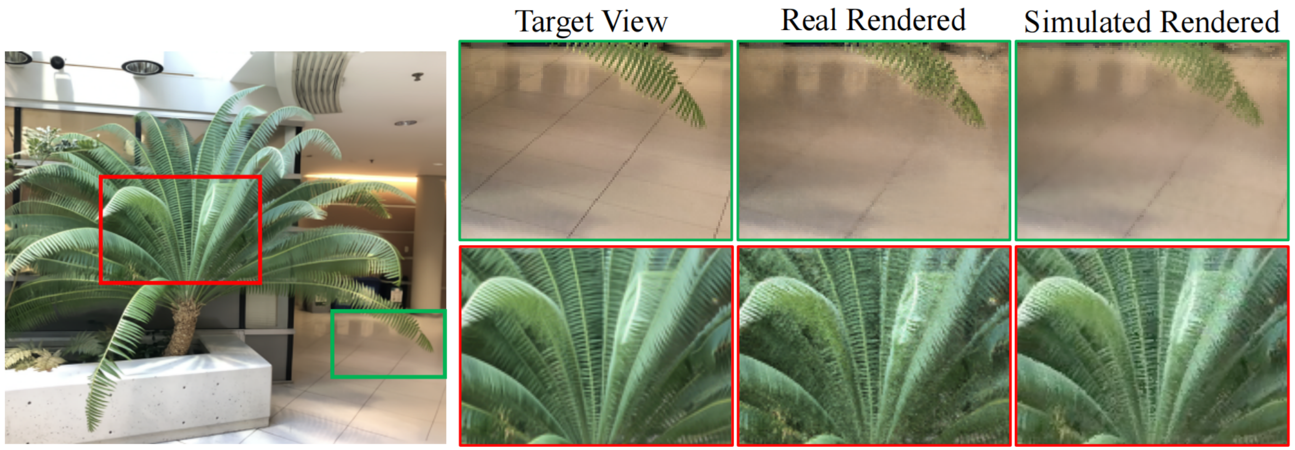
Some visualized region-adaptive masks

Method

NeRF-style degradation simulator (NDS)

- NDS pipeline





A visual example of real and simulated rendered views

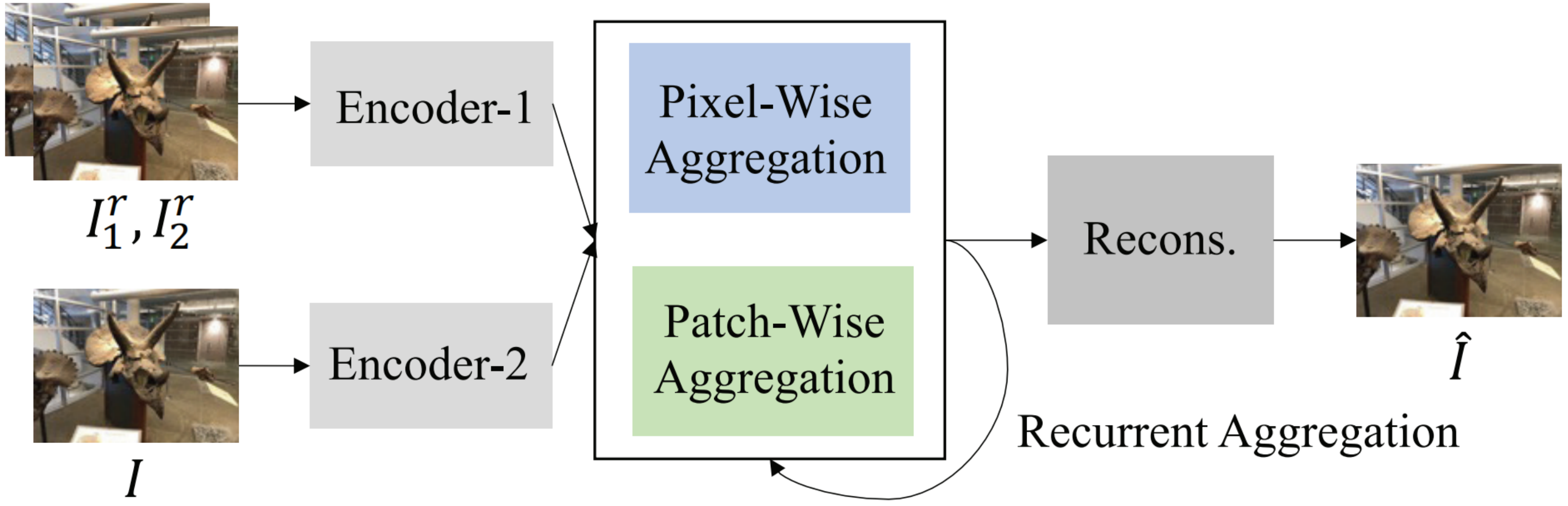
Method

Existing image/video restoration methods

- Image restoration approaches have limited enhancement ability solely dependent on the degraded views.
- Video restoration models cannot effectively handle the distinct viewpoint changes between an input view and its two reference views.

Inter-viewpoint mixer (IVM)

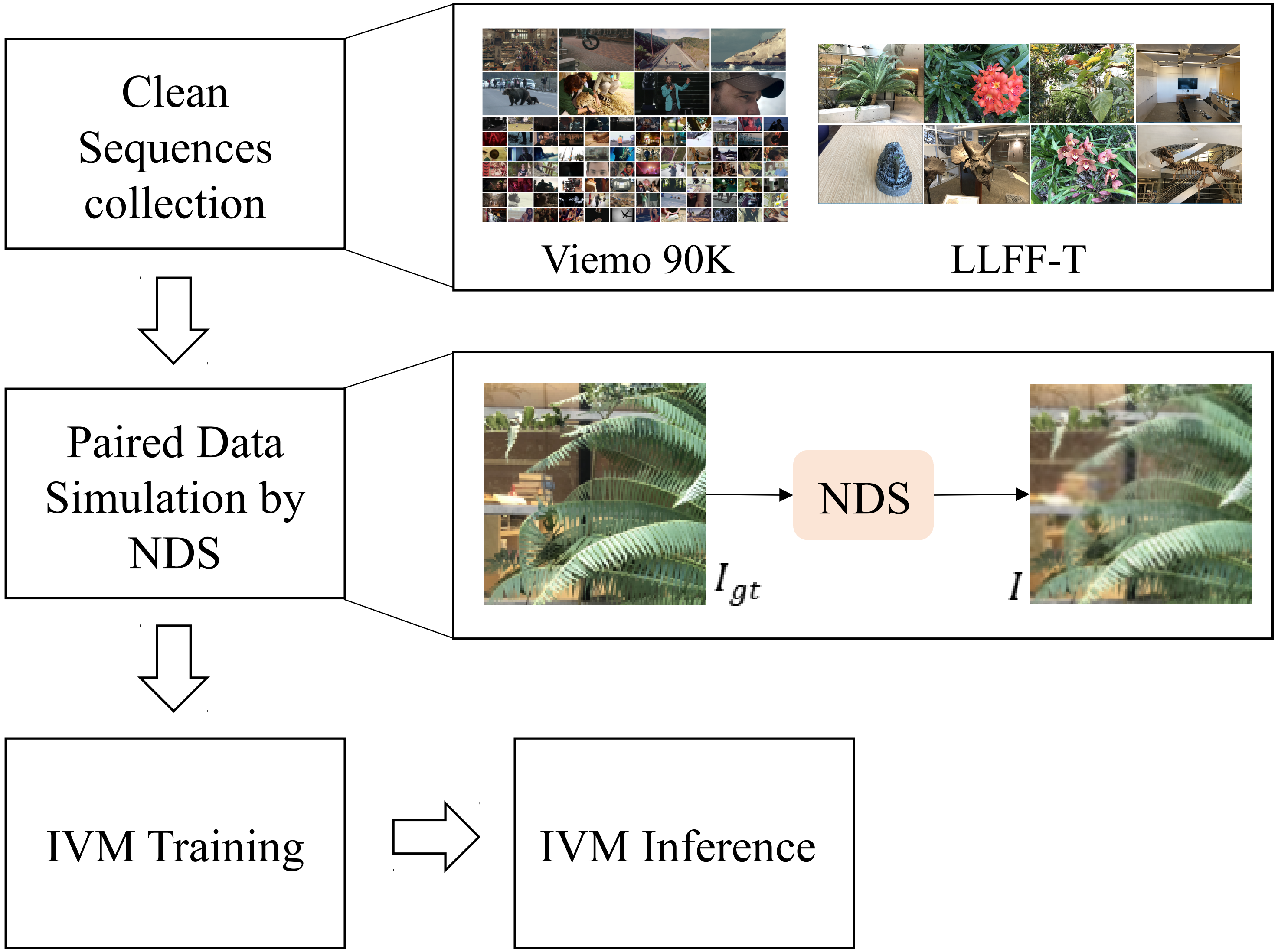
- hybrid pixel-wise and patch-wise aggregation to effective inter-viewpoint fusion..
- iterative aggregation to further improve the inter-viewpoint fusion accuracy.



Experiments

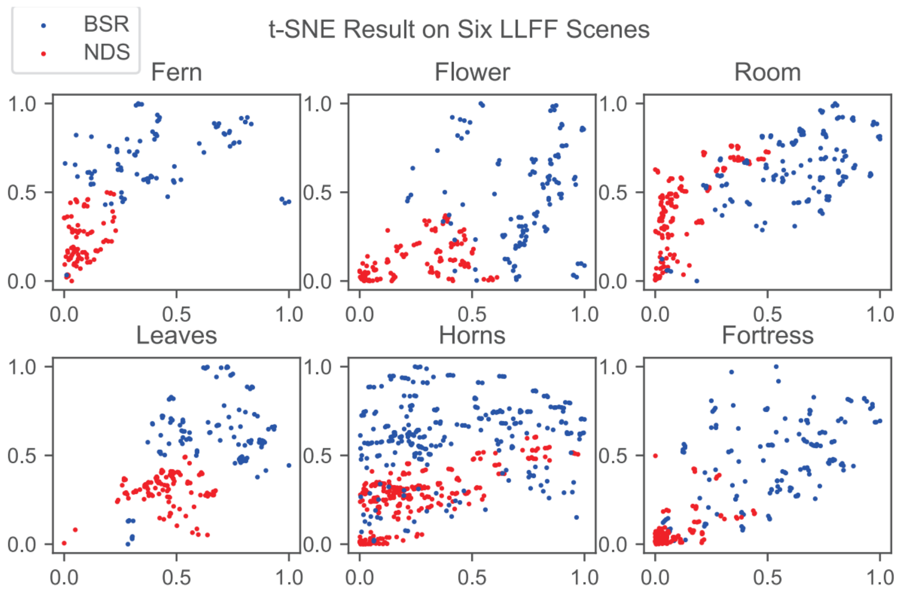
Datasets

- Training Sets:
  - Viemo 90K
  - LLFF-T(raing)
- Testing Sets:
  - LLFF
  - Tanks and Temples
  - Noisy LLFF Synthetic
- Evaluation Metrics:
  - PSNR
  - SSIM
  - LPISP



Ablation Study

(1) Statically evaluation of BSR and ours proposed NDS



Quantitative comparison between our NDS and BSR. 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.

(2) Impact of each presented degradation case in NDS

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 1. 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.

Ablation Study

(3) Quantitative comparison between BSR and NDS

Models	BSR	NDS	SwIR	IVM	PSNR	SSIM
$\mathrm{SwIR_B}$	✓		✓		26.20	0.834
$\mathrm{SwIR_N}$		✓	✓		**26.82**	**0.845**
$\mathrm{IVM_B}$	✓			✓	26.40	0.842
$\mathrm{IVM_N}$		✓		✓	**27.39**	**0.867**

Table 2. Comparison of our NDS and the BSR degradation models

(4) NDS enbales the resotration ability of existing image/video resotration methods for NeRF-render images.

Model	TensoRF(Base)	$\mathrm{SwIR_N}$	$\mathrm{DATSR_N}$	$\mathrm{EDVR_N}$	$\mathrm{VST_N}$
PSNR	26.70	26.82	26.84	26.88	26.79
SSIM	0.838	0.845	0.843	0.847	0.842

Table 3. Quantitative results of the improvements using existing image/video processing models trained on our simulated dataset

*We use subscripts ***N*****,** ***B***, to represent the models trained with our NDS dataset and BSR, respectively

Ablation Study

(5) Effects of our proposed IVM

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 4. Ablation studies of hybrid inter-viewpoint aggregation module. The running time is tested with an input size of 256×256

- Both the pixel-wise and patch-wise aggregations contribute to the final results.
- More iterations, higher performance.

(6) View selection strategy

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.

Our proposed view selection strategy can identify the two reference views that are most overlapped with the input (test) view.

Experiments

Quantitative analysis of our NeRFLiX

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**

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**

Quantitative results on the LLFF under LLFF-P2.

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**

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**

Improvement over TensoRF and Plenoxels on Noisy LLFF Synthetic

*We adopt 1008 × 756 resolution for LLFF-P1 and 504 × 376 resolution for LLFF-P2

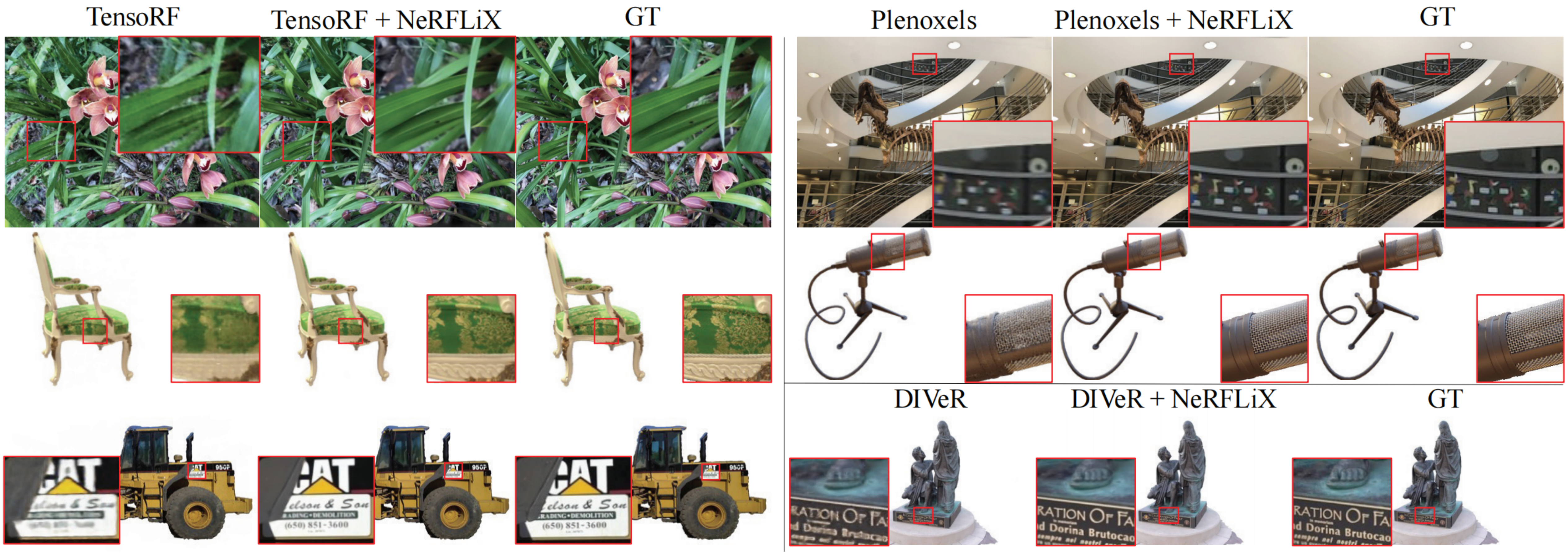
Experiments

Training Acceleration for NeRF Models

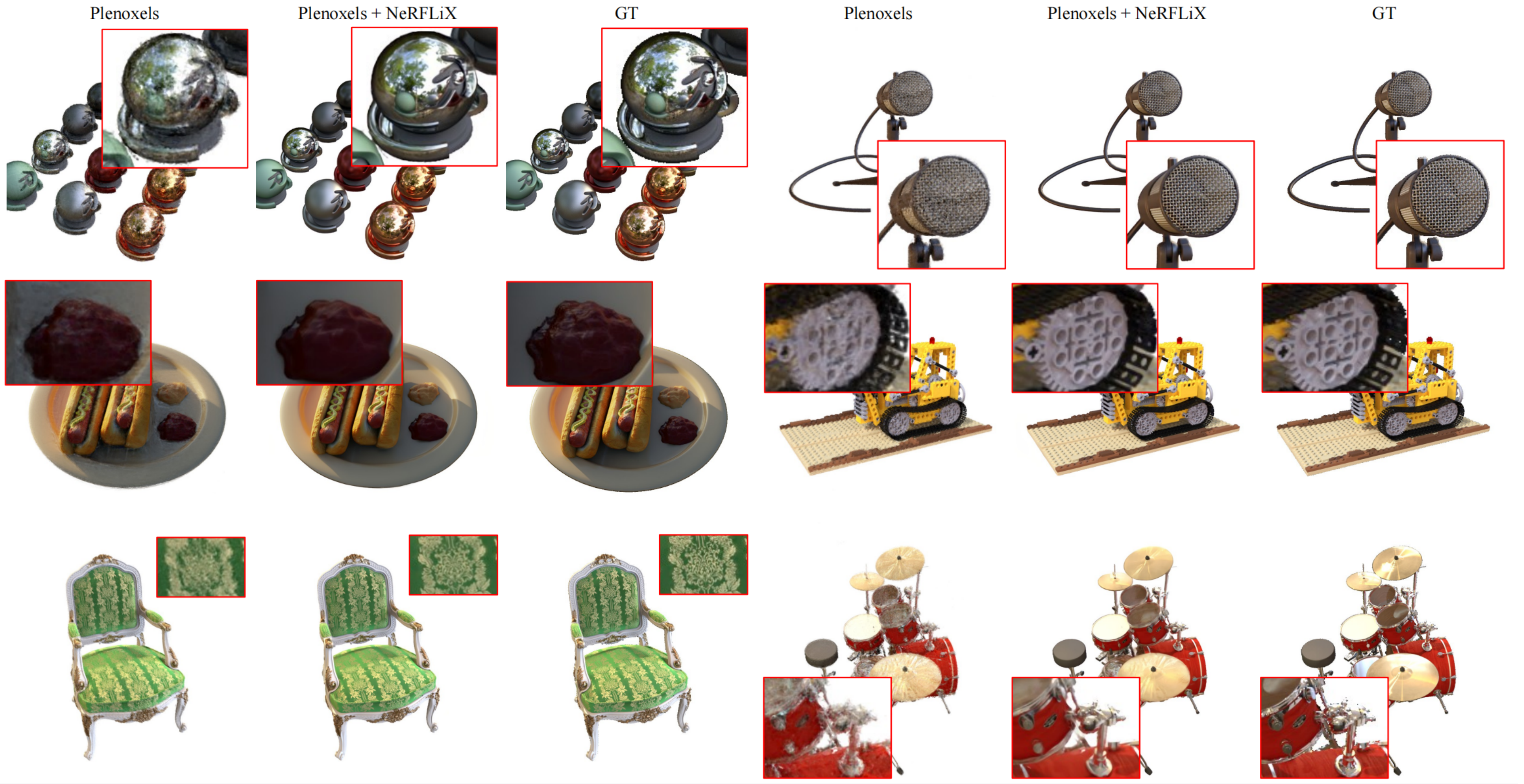
Method	PSNR (dB)↑/SSIM↑/LPIPS↓
TensoRF [7](4 hours)	26.73/ 0.839/ 0.204
TensoRF [7](**2 hours**)	26.18/ 0.819/ 0.230
[7](**2 hours**) + NeRFLiX	**27.14/ 0.858/ 0.165**
Plenoxels [16](24 minutes)	26.29/ 0.839/ 0.210
Plenoxels [16](**10 minutes**)	25.73/ 0.804/ 0.252
[16](**10 minutes**) + NeRFLiX	**26.60/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

Fewer training epoches, but better performance.

Visualization



Visualization



Visualization

Video Demo-3

TensoRF (ECCV'22)

TensoRF + NeRFLiX

Video Case 2(In-the-wild Scene 2)

Summary

NeRFLiX has the following advantages

- not require accurate camera poses.
- regardless of building complicated inverse rendering systems to regress object materials, and environment illumination.
- no re-training costs and can directly enhance all NeRF-rendered results significantly.
- Acclerate the training phases of SOTA NeRF models, while also enhancing the rendering quality.

Thank You!