NeRF-MS: Neural Radiance Fields with Multi-Sequence Supplementary Material

1. Controllable Experiment

We introduce the synthetic dataset based on NeRF lego dataset [3] and prove that our method is robust against the image count within a sequence. We show details of the synthetic dataset.

1.1. Synthetic Dataset

In the main text, we split the whole training views into different sequences $\{S_k\}_{k=1}^{M}$ and then apply appearance variation to each sequence. Specifically, fixing the total number of training views to 100, we reduce the image count per sequence from 20 to 2 by increasing the number of sequences M from 5 to 50. Thus, we explore the influence of different multi-sequence settings on our method. In supplementary, we split training views into 3 sequences with appearance variance for the comparison experiment.

Appearance Variation. To simulate the diverse appearance in different sequences, following [2, 1], we apply a random scale and shift transformation to the color values of images in each sequence as shown in Fig. 1. Specifically, for each image $I_i \in S_k$, we apply a transform $\widetilde{I}_{ij} = \max(0, \min(1, s_{kj} I_{ij} + b_{kj}))$ for each RGB channels j, where scale $s_{kj} \sim \mathcal{U}(0.8, 1.2)$ and shift $b_{kj} \sim \mathcal{U}(-0.2, 0.2)$ are sampled randomly for each sequence S_k and RGB channel j. We apply the color perturbations to all training sequences except the first sequence, since we generate test views for evaluation with the appearance code of the first image of the first sequence.

1.2. Comparison with State-of-art Method

Quantitative results on synthetic dataset with appearance variation are shown in Table. 1. With multi-sequence training data, our method outperforms the baselines and approximates to NeRF trained on original lego dataset. Fig. 2 shows the qualitative results of the comparison experiment. NeRF-W and Ha-NeRF are unable to disentangle color perturbations in multi-sequence training data, leading to color bias in novel view synthesis. The last row in Fig. 2 illustrates that our method can reconstruct the 3D consistent reflections in multi-sequence tasks.

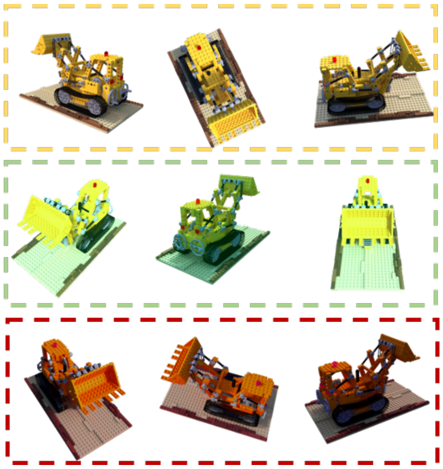

Figure 1: **Examples of perturbations applied to the synthetic dataset.** We show 3 example sequences with appearance variation.

	Dataset	↑PSNR	↑SSIM	↓LPIPS
NeRF	Original	32.21	0.9566	0.0167
NeRF	Appearance Variation	18.38	0.8511	0.1252
NeRF-W		29.64	0.9449	0.0349
Ha-NeRF		30.56	0.9525	0.0318
Ours		**31.74**	**0.9566**	**0.0259**
NeRF-W	Sequences w/ Same Appearance	29.03	0.9487	0.0308
Ha-NeRF		26.96	0.9421	0.0404
Ours		**32.68**	**0.9656**	**0.0193**

Table 1: **Quantitative results on the synthetic dataset.** Original dataset refers to the synthetic lego dataset without any perturbation.

1.3. Robustness against Sequences with Same Appearance

Although the triplet loss regularizes the distance of latent codes from different sequences, our method is robust while meeting similar sequences without appearance variation. We split the original lego dataset into 3 sequences, which have the same appearance, and evaluate NeRF-MS and baselines on such dataset. Quantitative results are shown in the last 3 rows of Table. 1 and our method out-

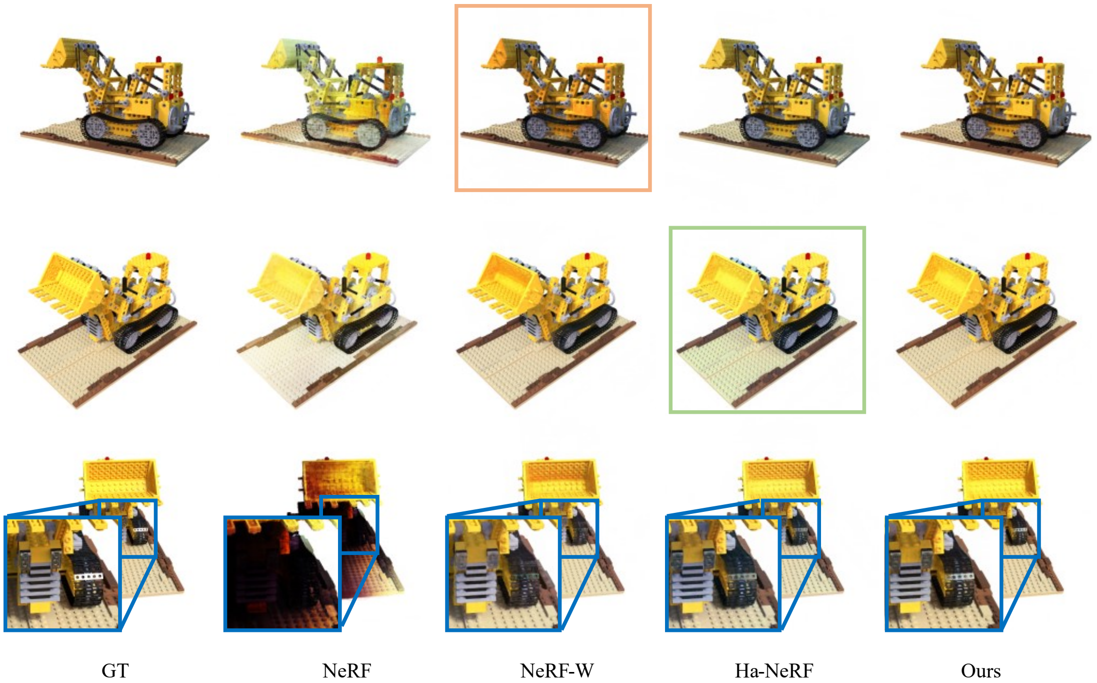

Figure 2: **Qualitative results on the synthetic dataset.** Our method is able to reconstruct reflections, high-frequency geometry details and disentangle appearance variation. The figures framed in green and orange illustrate that NeRF-W and Ha-NeRF suffer from color shifts in multi-sequence tasks.

performs other baselines since triplet loss reduces the flexibility of latent code and prevents overfitting.

2. Additional Qualitative Results

2.1. Visualization of Transient Decomposition

We present predicted uncertainty maps and composition results of our method and baselines in Fig. 3. Due to our transient decomposition module, NeRF-MS predicts less uncertainty in static areas and has better reconstruction for both static and transient objects.

2.2. Qualitative Results on NeRF-OSR dataset

To demonstrate the effectiveness of our method, we show more qualitative results of test set on NeRF-OSR dataset. Our method can reconstruct more high-frequency details, as shown in example 1, 3 and 4 in Fig. 4. Besides, NeRF-W suffers from color bias (see example 1 and 4) and floaters (see example 2). Due to the high flexibility of the 2D transient masks predictor, Ha-NeRF tends to decompose objects from static scenes excessively (see example 1 and 2).

2.3. T-SNE Visualization

We show the T-SNE visualization results of appearance codes from different methods in Fig. 5. The distribution of appearance codes from NeRF-W and Ha-NeRF is messy since appearance codes from different sequences are overlapped due to overfitting. However, our method can build up better appearance latent space where the appearance codes within a sequence bunch up.

3. Performance with Larger Regularization

We find that with larger regularization weight $\lambda = 1$ and smaller margin $m = 1$ of triplet loss, the NeRF-MS can achieve better performance under most scenes as shown in Fig. 2. We encourage readers to use these new hyperparameters to get better results than the original paper.

References

[1] Xingyu Chen, Qi Zhang, Xiaoyu Li, Yue Chen, Ying Feng, Xuan Wang, and Jue Wang. Hallucinated neural radiance fields in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12943–12952, 2022. 1

[2] 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, 4

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

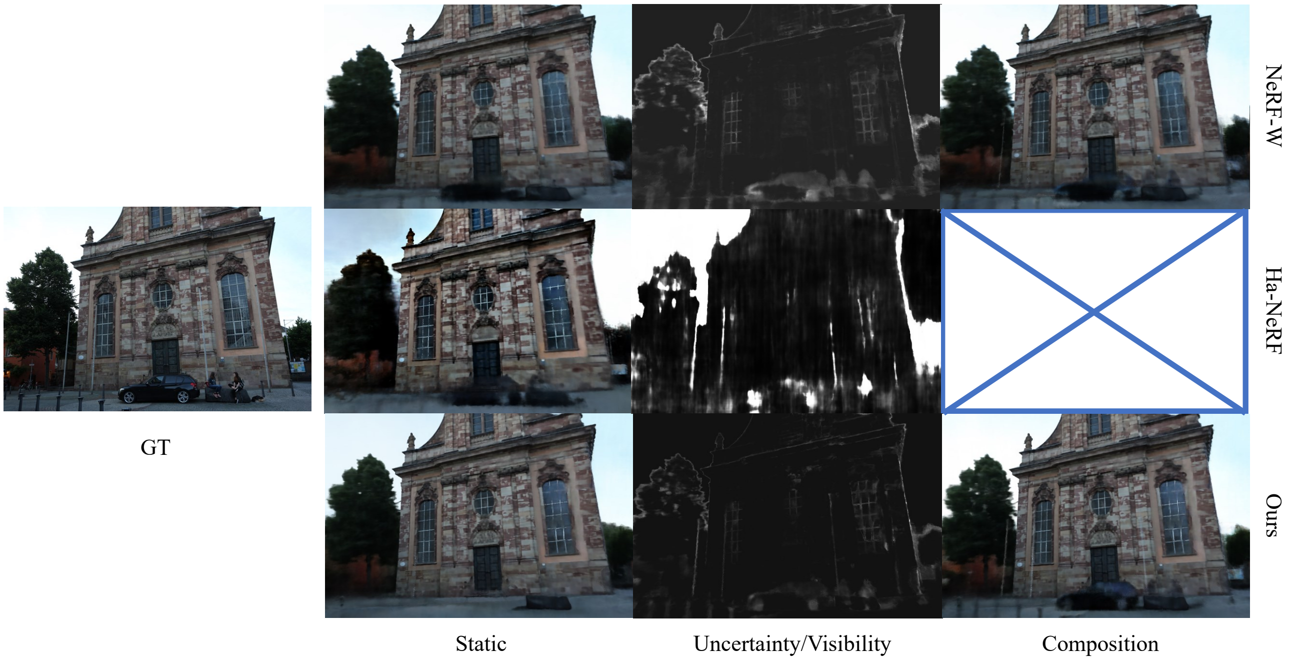

Figure 3: **Additional qualitative results of Transient Decomposition.** Note that Ha-NeRF predicts visibility map instead of uncertainty map and it can't render the composition result.

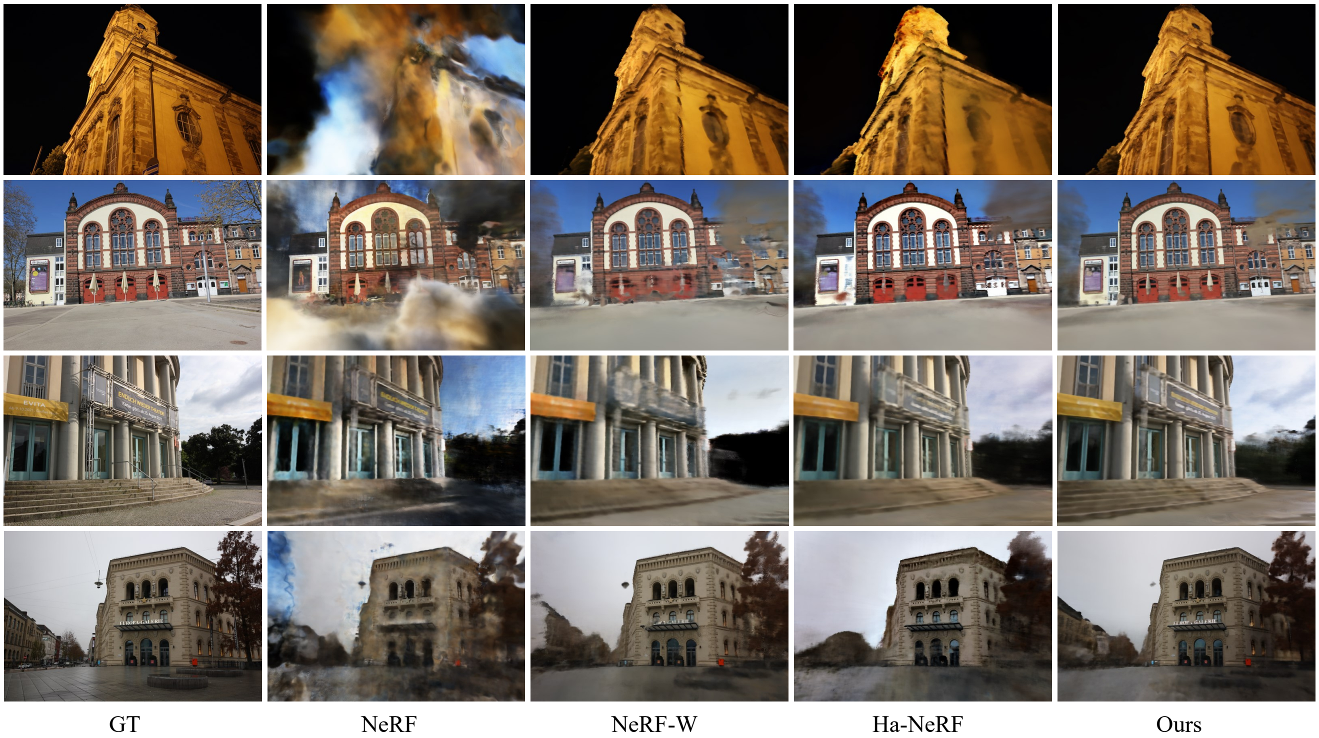

Figure 4: **Additional qualitative results on NeRF-OSR dataset.**

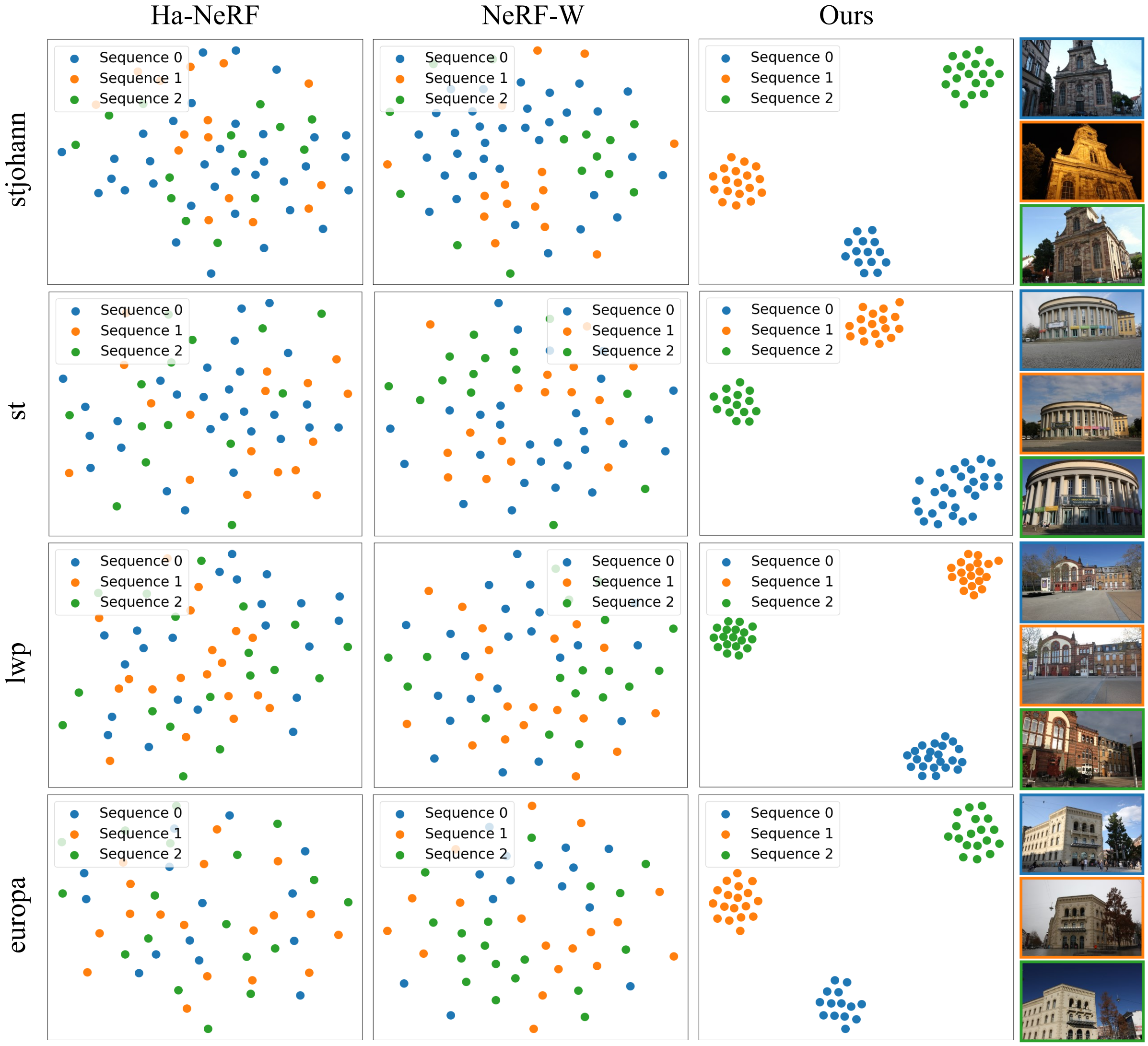

Figure 5: **T-SNE visualization of appearance codes for each scene.**

	stjohann			lwp			st			europa		
	↑PSNR	↑SSIM	↓LPIPS	↑PSNR	↑SSIM	↓LPIPS	↑PSNR	↑SSIM	↓LPIPS	↑PSNR	↑SSIM	↓LPIPS
NeRF-W [2]	21.230	0.6673	0.4255	19.607	0.6159	0.4453	20.310	0.6067	0.4380	20.000	0.6899	0.3397
Ours	**22.839**	**0.7933**	**0.2347**	21.904	0.7187	0.3361	20.675	0.6304	0.4021	21.028	0.7214	**0.2939**
Ours*	22.292	0.7844	0.2469	**22.450**	**0.7424**	**0.3001**	**22.281**	**0.6453**	**0.3979**	**21.461**	**0.7282**	0.3119

Table 2: **Quantitative results with different hyperparameter.** Ours* shows that, with $\lambda = 1, m = 1$, our method can achieve better PSNR, SSIM and competitive LPIPS.