Supplementary Material for "Hallucinated Neural Radiance Fields in the Wild"

Xingyu Chen[1*] Qi Zhang[2†] Xiaoyu Li[2] Yue Chen[1]
Ying Feng[2] Xuan Wang[2] Jue Wang[2]
[1]Xi'an Jiaotong University [2]Tencent AI Lab

A. Ray Sampling Strategy

To improve efficiency, we sample a grid of rays and combine them as the image $\mathcal{I}_i^r$ instead of rendering a whole image during training. [64]. The sampled patch is determined by a square 2D coordinates with random center followed by cropping of images. We use the cropping size of 0.8, and sample 32×32 rays at intervals during training, which means we render out a low resolution image to feed into the consistency loss. Thanks to the high cropping size of 0.8, the sampling points almost cover the whole image. We assume the appearance of image cross different resolution is consistent, which concludes that the patch contains the same global appearance of the whole image after sampling.

B. Experiments

We encourage the reader to view the video results included in the supplementary materials for an intuitive hallucinated experience of visiting cultural landmarks. Here we provide more evaluations of our hallucinated neural radiance fields (Ha-NeRF) and the state-of-the-art NeRF in the wild method (NeRF-W) [63].

B.1. Synthetic Dataset

The components of Ha-NeRF are designed to deal with specific problems, *i.e.*, color hallucination and anti-occlusion. Unfortunately, the uncontrolled captures of the Phototourism dataset make it challenging to demonstrate the effectiveness of each component. For this reason, we present an ablation study in which we construct variations (*e.g.*, color and occlusion) of a synthetic dataset used in [63]. We manually introduce the phenomena we expect to find in in-the-wild imagery.

For fair comparisons, the inputs of NeRF-W are the same as Ha-NeRF, and thus the appearances of test images are obtained from another image with the same appearance in the training data. Tab. 2 and Fig. 1 shows the quantitative and qualitative results of ablation study about color hallucination and anti-occlusion. The proposed method outperforms the baselines. Specifically, for anti-occlusion, our method tackles the occlusions very well, which verifies the accuracy of our image-dependent occlusion. For color hallucination, the performance of the proposed method is better than baselines, especially on the color shifts and highlight regions with high-frequency information, as shown in Fig. 1. For the experiment on the combination of color and occlusion, the performance of the proposed method shows that our method could decompose the occlusion and appearance well compared with NeRF-W. All of the results demonstrate the superior performance of the proposed method.

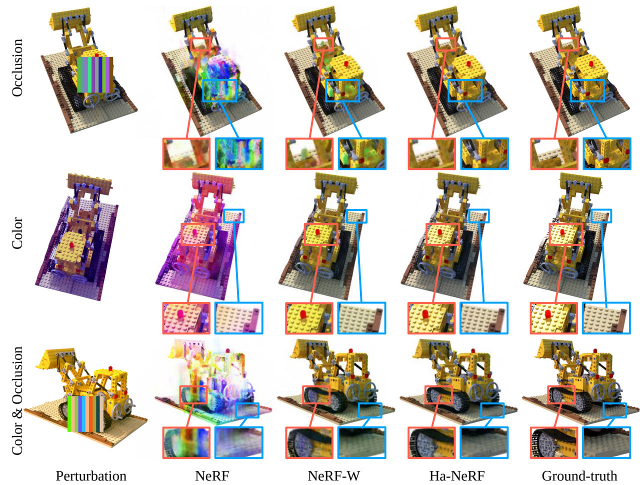

Figure 1. Example dataset perturbations and rendering from NeRF, NeRF-W, Ha-NeRF and ground-truth for fair comparisons.

B.2. View-consistent Hallucination.

As the camera moves, renderings of NeRF-W conditioned on the same appearance embedding appear to have an inconsistent appearance. On the contrary, we can perform view-consistent appearance rendering thanks to our view-consistent loss. Because the aligned image pairs are challenging to collect (tuples of images captured in different views under the same appearance conditions), we use the synthetic dataset [63] with ground truth and manually introduce color perturbation for training. The left of Fig. 2

[1]Work done during an internship at Tencent AI Lab.
[2]Corresponding Author.

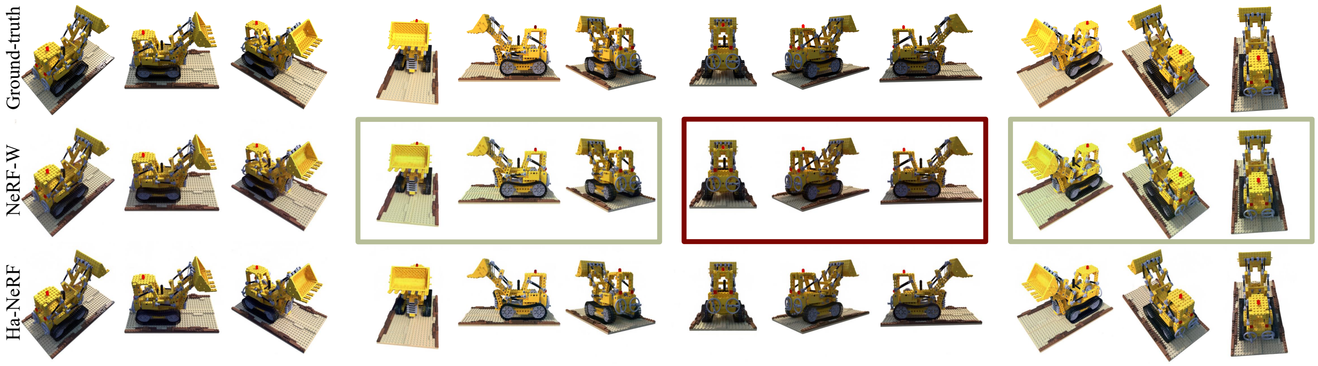

Figure 2. Results of the experiment on synthetic Lego dataset with ground truth and manually introduce color perturbation for training. Figure framed in green and red shows that the appearance of images rendered by NeRF-W has green bias and red bias with respect to ground truth.

	1e-3	2e-3	3e-3	4e-3	5e-3	6e-3	7e-3	8e-3	9e-3	1e-2	2e-2	3e-2
Brandenburg Gate	\	22.55	23.62	23.46	24.01	24.04	**24.36**	23.83	23.36	23.79	23.26	\
Sacre Coeur	\	\	19.97	19.45	19.78	20.02	20.24	**20.43**	20.01	19.77	19.75	\

Table 1. PSNR with respect to regularization weight.

	Dataset	PSNR ↑	SSIM ↑	LPIPS ↓
NeRF	Original	32.23	.9566	.0167
NeRF	Color Perturbation	20.89	.9006	.1195
NeRF-W		28.85	.9370	.0337
Ha-NeRF		29.60	.9375	.0294
NeRF	Occlusion	19.04	.8627	.1394
NeRF-W		28.12	.9336	.0434
Ha-NeRF		30.93	.9494	.0233
NeRF	Color Perturbation & Occlusion	17.17	.8385	.1959
NeRF-W		27.10	.9211	.0525
Ha-NeRF		28.41	.9349	.0360

Table 2. Quantitative results from experiments on the synthetic dataset in different cases. Ha-NeRF outperforms the others on all evaluation metrics.

shows that both NeRF-W and Ha-NeRF can render images with the correct appearance. However, the right of Fig. 2 framed in green and red indicate that the appearance of images rendered by NeRF-W has green bias and red bias with respect to ground truth. The experimental results demonstrate that NeRF-W suffers from inconsistent appearance when the camera moves, while Ha-NeRF is consistently consistent with ground truth.

B.3. Ablation on the regularization.

The regularization weight λ_o of invisible probability is crucial to the performance of our method. We set the regularization weight λ_o to 6×10^{-3} as a empirical parameter in all experiments. Here we add an ablation experiment to discuss the effects of different λ_o. As shown in the Tab. 1, too small weight leads the model to ignore static phenomena, while too large weight leads the model to construct ghosting artifacts from transient objects. Instead of using a fixed weight for all experiments, floating the weight between 2×10^{-3} and 2×10^{-2} may achieve preferable results for specific scenes.

B.4. Appearance Hallucination.

We render more images in Fig. 3 produced by Ha-NeRF using different appearance vectors extracted from example images. We also show the results of NeRF-W where appearance vectors are optimized during training. Notice that Ha-NeRF hallucinates realistic images while NeRF-W suffers from global color shifts (such as weather, season and postprocessing filters) compared with the example images. Moreover, Fig. 4 shows that Ha-NeRF can capture the high-frequency information of appearance and hallucinate the sunshine and colored light reflection of the scene.

B.5. Appearance Hallucination Cross Datasets

We perform more example-guided appearance transfer by a user-provided example image from a different dataset. As shown in Fig. 5, we can hallucinate new appearance for "Trevi Fountain" condition on the example image of "Brandenburg Gate", vice versa (Fig. 6).

References

[63] 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 *IEEE CVPR*, pages 7210–7219, 2021. 1

[64] Katja Schwarz, Yiyi Liao, Michael Niemeyer, and Andreas Geiger. Graf: Generative radiance fields for 3d-aware image synthesis. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 1

Figure 3. Hallucination in the “Brandenburg Gate” dataset with the global color shifts, such as weather, season and postprocessing filters. There are the images whose viewing direction is the same as the leftmost column content images, and the appearance is conditioned on the top line example appearance images.

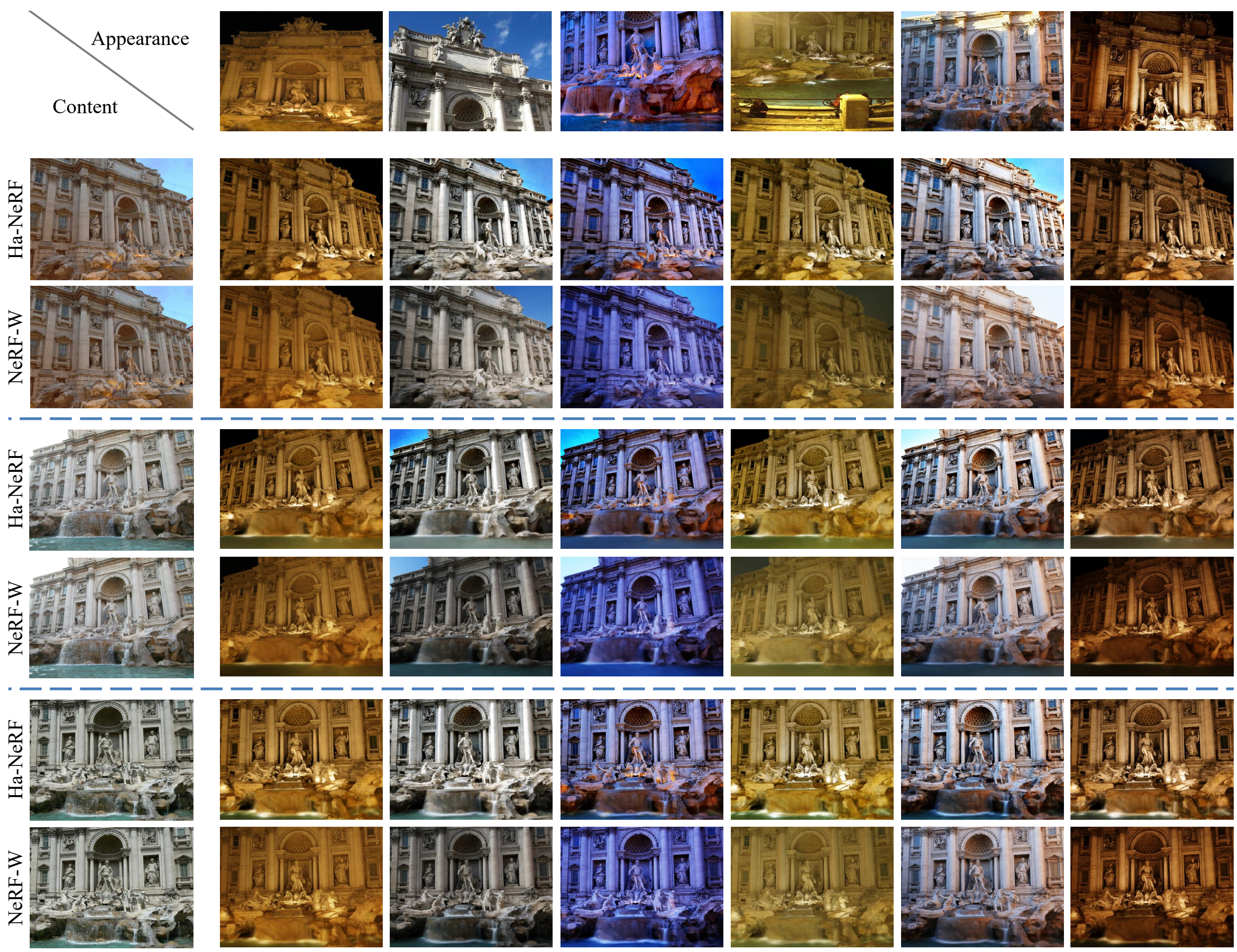

Figure 4. Hallucination in the "Trevi Fountain" dataset with high-frequency information of appearance, such as sunshine and colored light reflection. There are the images whose viewing direction is the same as the leftmost column content images, and the appearance is conditioned on the top line example appearance images.

Figure 5. Cross dataset hallucination in the “Trevi Fountain” condition on the example image of “Brandenburg Gate” with the global color shifts, such as weather, season, and postprocessing filters. There are the images whose content is the same as the leftmost column content images, and the appearance is conditioned on the top line example appearance images of another dataset.

Figure 6. Cross dataset hallucination in the “Brandenburg Gate” condition on the example image of “Trevi Fountain” with the global color shifts, such as weather, season, and postprocessing filters. There are the images whose content is the same as the leftmost column content images, and the appearance is conditioned on the top line example appearance images of another dataset.