Supplementary Material: Non-uniform Sampling Strategies for NeRF on 360° images

Takashi Otonari[1]
otonari@hal.t.u-tokyo.ac.jp
Satoshi Ikehata[2]
sikehata@nii.ac.jp
Kiyoharu Aizawa[1]
aizawa@hal.t.u-tokyo.ac.jp

[1] The University of Tokyo
Tokyo, Japan
[2] National Institute of Informatics
Tokyo, Japan

1 More Results with DietNeRF, AugNeRF and NeRF++ on Synth360

1.1 Details of Methods

In the main paper, we demonstrated the effectiveness of our non-uniform sampling strategies in different NeRF-like models. We have implemented the proposed method in the authors' publicly available source codes with minimal changes. Here we briefly describe the details of individual algorithms.

DietNeRF [4] added an auxiliary semantic consistency loss to the näive NeRF. The weight of an auxiliary semantic consistency loss was set to 0.01 and calculated by using an image that was resized to 224×224 for every 10 iterations.

AugNeRF [7] brought the power of robust data augmentations into regularizing the NeRF training. Augmentation was given to the intermediate feature, pre-rendering output, and input coordinate levels. The perturbations were estimated by multi-step Projected Gradient Descent. The weights of photometric loss and adversarial loss were set to 0.5.

NeRF++ [6] was specially optimized for unbounded scenes by explicitly decomposing an entire scene into the foreground and background ones with different distance divisions. For this decomposition, NeRF++ normalizes each scene so that all cameras were inside the sphere of radius $\frac{1}{1.2}$ and the average camera position of all the training cameras was the center of the sphere.

1.2 PSNR/SSIM Curves

In the main paper, we showed the PSNR/SSIM scores at 100,000-th iteration. Here, we also show PSNR/SSIM curves averaged over indoor/outdoor datasets while training DietNeRF [4], AugNeRF [7], and NeRF++ [6] w/ or w/o our non-uniform sampling strategies.

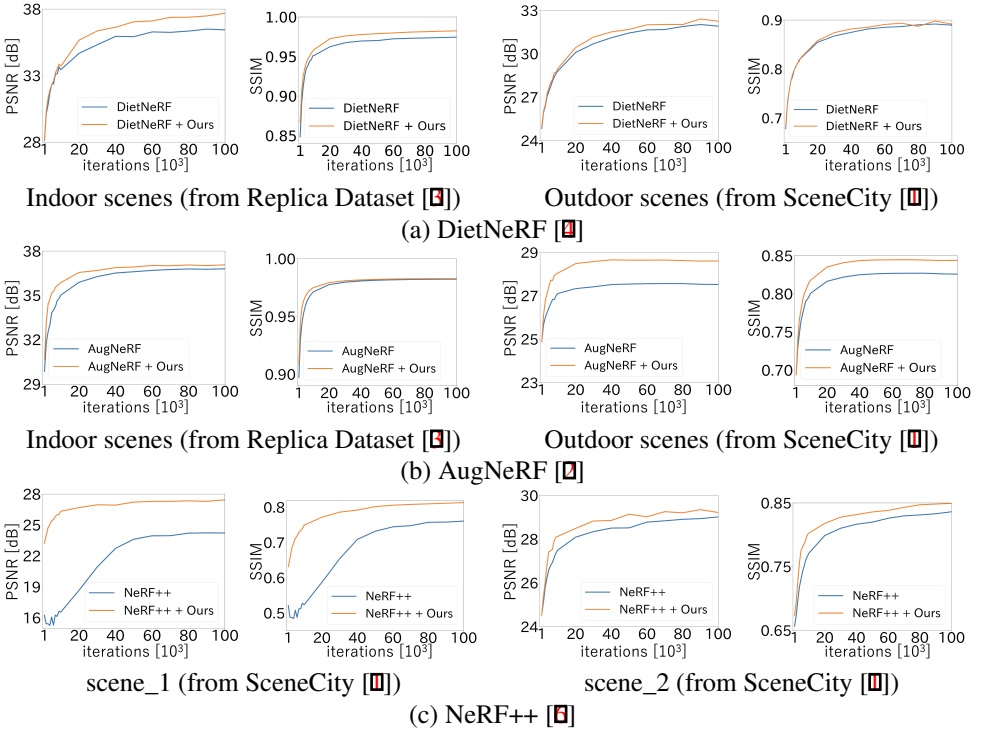


Figure 1: PSNR/SSIM curves of various advanced variants of NeRF combined with our proposed method.

The results are shown in Fig. 1. In summary, we observe the consistent improvement of learning curves with our method which shows both more efficient training and better accuracy at convergence.

As mentioned, we found that NeRF++ often showed significant performance degradation when the foreground and background were not well separated (See Fig. 2-top). Specifically, NeRF++ failed to be trained on indoor scenes in our Synth360 (*i.e.*, Replica Dataset); therefore, we only show the result on outdoor scenes (*i.e.*, SceneCity; scene_1 and scene_2). It is interesting to see that the näive NeRF++ also failed to separate the foreground and background of the indoor dataset. However, combined with the proposed method, the learning of NeRF++ progressed appropriately on them. The characteristics of the individual algorithms are outside the scope of this work and will not be discussed further (Fig. 2-bottom).

1.3 Qualitative Results

The qualitative results of DietNeRF [4], AugNeRF [2], and NeRF++ [6] w/ and w/o our non-uniform sampling strategies are shown in Fig. 3. We also observe our proposed method is also effective for advanced variants of NeRF: better high-frequency texture recovery and less artifacts.

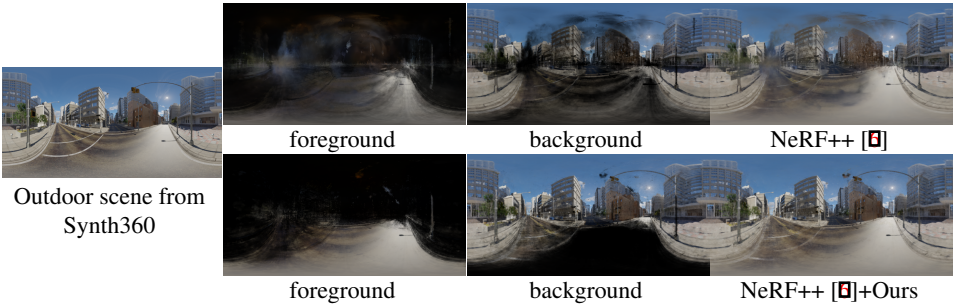


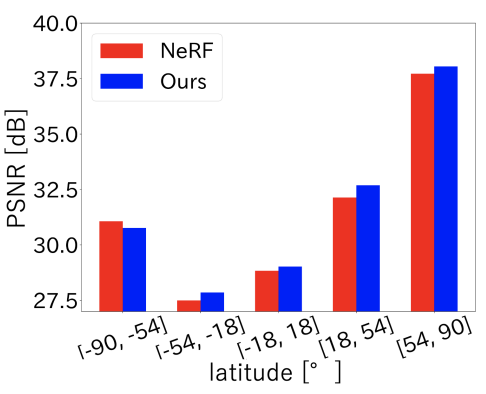
Figure 2: Unsuccessful case of NeRF++ [6]. We found that NeRF++ sometimes fails to separate between foreground and background. However, combining our proposed method often improved this failure case.

Figure 3: Qualitative comparison using DietNeRF [4], AugNeRF [7], and NeRF++ [6] on Synth360.

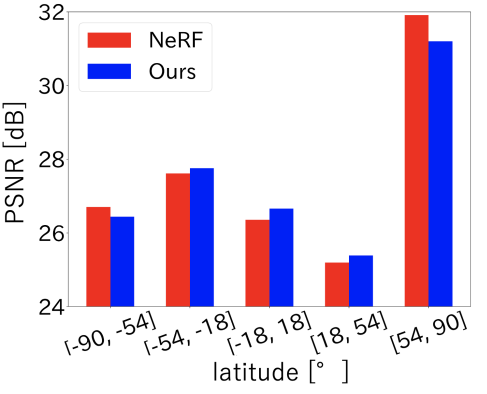
2 Detailed Behavior of Non-Uniform Sampling Strategies

So far, we have evaluated our method based on quantitative and qualitative comparisons on an entire image and validated that our non-uniform sampling strategies are effective for improving both learning curves and the reconstruction quality at the convergence despite its simplicity. However, since the motivation of the proposed method is to dynamically change the sampling probability according to the amount of sphere-to-plane projection distortion at high latitudes and the content of 360° images with a wide field of view, this section verifies whether our strategies actually work as intended.

First, to validate the behavior of our distortion-aware ray sampling, we divided the ERP coordinate into five regions by latitude and made quantitative comparisons in each region as



Indoor scenes (from Replica Dataset [3])



Outdoor scenes (from SceneCity [1])

Figure 4: Comparison by PSNR for each region of latitude. Latitude is assigned as -90° at the bottom of the image and 90° at the top of the image. Thus, [-90, -54] and [54, 90] are high latitude regions and [-18, 18] are low latitude regions.

Table 1: Quantitative comparison (PSNR) using NeRF [5] and Ours on low-frequency texture and high-frequency texture regions. The best is highlighted.

	Indoor scenes (Synth360)		Outdoor scenes (Synth360)	
Method	low-frequency	high-frequency	low-frequency	high-frequency
NeRF [5]	44.63	29.68	**38.28**	23.51
Ours	**44.85**	**31.63**	37.42	**24.86**

shown in Fig. 4. As expected, since our distortion-aware sampling strategy assigns a higher sampling probability in lower-latitude regions, the reconstruction accuracy in low-latitude regions improved while one in high-latitude regions often slightly degraded. It is important to note that PSNR is computed in ERP coordinates; therefore, the reconstruction accuracy at high latitudes, where the amount of information is far less than at low latitudes, is less important for most practical applications.

Second, to validate the behavior of our content-aware ray sampling, we cropped high-frequency and low-frequency texture crops of 60×60 and evaluated our method on each crop individually. We observed that our content-aware ray sampling probability assigns large probability values to regions with many edges. Therefore, we considered regions with many edges to be high-frequency regions and regions with few edges to be low-frequency regions. Specifically, a 3×3 laplacian filter was applied to the ERP image, and crops with large outputs were defined as high-frequency texture regions, while regions with small values were defined as low-frequency texture regions. One low-frequency texture region and one high-frequency texture region were extracted for each test image. As shown in Table 1, our proposed method fairly improved the reconstruction accuracy for the high-frequency texture crops because our content-aware ray sampling strategy increased the number of samples in challenging high-frequency texture regions. On the other hand, though the proposed method reduced the number of samples for the low-frequency crops, which are easy to learn, its accuracy degradation was found to be small.

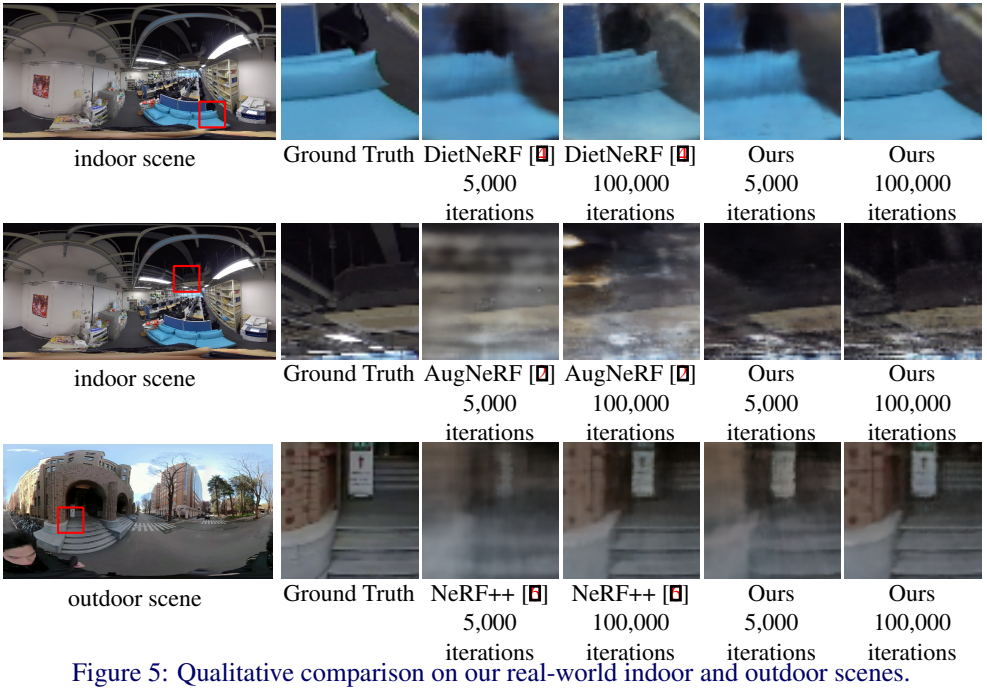


Figure 5: Qualitative comparison on our real-world indoor and outdoor scenes.

3 Qualitative comparison using advanced variants of NeRF on real-world scenes

The qualitative results of DietNeRF [4], AugNeRF [2], and NeRF++ [6] combined with our proposed method on real-world indoor and outdoor scenes are shown in Fig. 5. We found that our proposed method works well for real-world scenes and improves the synthesis quality in the high-frequency texture regions.

References

[1] SceneCity. https://www.cgchan.com/store/scenecity.

[2] Tianlong Chen, Peihao Wang, Zhiwen Fan, and Zhangyang Wang. Aug-nerf: Training stronger neural radiance fields with triple-level physically-grounded augmentations. In *CVPR*, 2022.

[3] T. Whelan J. Straub, L. Ma, Y. Chen, E. Wijmans, S. Green, J. J. Engel, R. Mur-Artal, C. Ren, S. Verma, A. Clarkson, M. Yan, B. Budge, Y. Yan, X. Pan, J. Yon, Y. Zou, K. Leon, N. Carter, J. Briales, T. Gillingham, E. Mueggler, L. Pesqueira, M. Savva, D. Batra, H. M. Strasdat, R. D. Nardi, M. Goesele, S. Lovegrove, and R. Newcombe. The Replica dataset: A digital replica of indoor spaces. *arXiv:1906.05797*, 2019.

[4] A. Jain, M. Tancik, and P. Abbeel. Putting nerf on a diet: Semantically consistent few-shot view synthesis. In *ICCV*, 2021.

[5] 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. In *ECCV*, 2020.

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