Supplementary Material
Neural Plenoptic Sampling: Learning Light-field from Thousands of Imaginary Eyes

Junxuan Li, Yujiao Shi, and Hongdong Li

Australian National University
{junxuan.li,yujiao.shi,hongdong.li}@anu.edu.au

1 Comparison with Conventional NVS Approaches on 360° Panorama Synthesis

In this section, we first review recent works on 360° panorama synthesis. Zheng *et al.* [10] propose a representation, named layered depth panorama (LDP), to create a layered representation with a full field of view from a sparse set of images taken by a hand-held camera. They focus on a scene representation method instead of synthesizing new views. Bertel *et al.* [2] investigate two blending methods for interpolating novel views from two nearby views, one is a linear blending, and the other is a view-dependent flow-based blending. However, both blending methods require that the position of novel views should be in the trajectory of input views. Serrano *et al.* [7] propose to synthesize new views from a fixed viewpoint 360° video. In contrast to the above methods, we aim to synthesize free-viewpoint panoramas from a few unstructured and sparse input panoramic images.

Huang *et al.* [4] employ a typical depth-warp-refine procedure in synthesizing new views. They estimate the depth map for each input image and reconstruct the 3D point cloud by finding correspondences between input images using hand-crafted features. They then synthesize new views from the reconstructed point cloud. The code of their work is not released. Thus we cannot fairly compare with them. Instead, we employ a more advanced deep-based method [8] to estimate depth maps for input images.

The qualitative comparison results are presented in Fig. 1, respectively. From the figure, it can be seen that there are severe distortions in the synthesized images by typical depth-warp-refine (*i.e.*, 360SD-Net [8]) strategy, while the synthesized images by our method are much similar to the ground truth. The numerical evaluations in the main paper also demonstrate that our method significantly outperforms the conventional depth-warp-refine procedure in synthesizing new views.

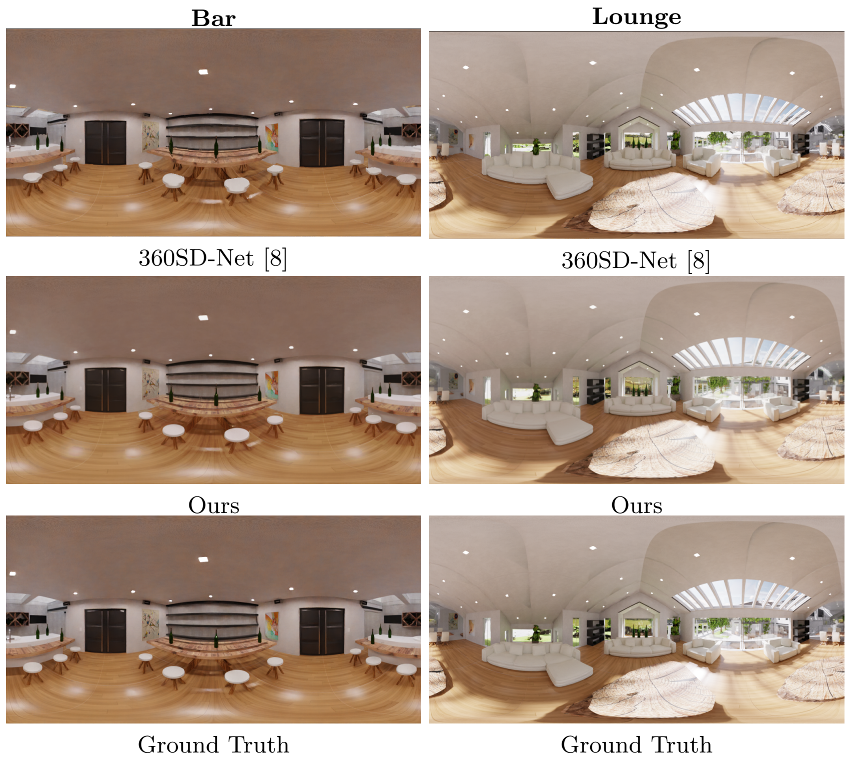

Fig. 1: Qualitative comparison of our algorithm with the conventional depth-warp-refine strategy (*i.e.*, 360SD-Net [8]). Best viewed in color on screen.

2 Experiments on 360SD-Net dataset and Supplementary Video

As mentioned in the paper, the performance on 360SD-Net [8] dataset is hard to be quantitatively evaluated due to the lack of ground truth data. Hence, we qualitatively visualize the synthesized images by our method in the supplementary video. Furthermore, we synthesize continuously generated images by our method on the four synthetic scenes when roaming around the space. Please refer to our supplementary video for the results.

3 Comparison With NeRF and NeRF Variant

We also present addition results on comparison with NeRF [5] and its variant NeRF++ [9]. In this comparison, all of the methods take eight unstructured views as input. The qualitative comparison results are presented in Fig. 2 , Fig. 3, and Fig. 4. It can be seen that our method achieves similar performance with NeRF and NeRF++, with marginal improvements in most of the scenarios.

Fig. 2: Qualitative comparison with NeRF and NeRF++ on our generated scenes "Lounge".

NeRF and NeRF++ aim to estimate the radiance emitted by scene points at any position and direction, while our method is designed to recover the irradiance perceived by an observer from any point and direction. In essence, our formulation is more close to the plenoptic sampling invented by Adelson and Bergen [1] Since NeRF and Nerf++ need to sample points along viewing rays and render them in a back-to-front order, they require hundreds of network calls when synthesizing an image. Thus their rendering time is very long. In contrast, our method directly outputs the color information given a viewing ray. Thus, our training and testing time are relatively shorter.

A recent work, DoNeRF [6], shares some similarity with ours. Both DoNeRF and our method first regress the depth for a target viewing ray. The difference is that DoNeRF has the ground-truth depth map for each viewing ray during

Fig. 3: Qualitative comparison with NeRF and NeRF++ on our generated scenes "Livingroom".

training, while our approach offers a self-supervision for the target view depth regression. FastNeRF [3] is another recent work that is proposed to accelerate the rendering speed during inference. Their approach changes the network architecture and explores a way to cache a number of pre-sampled scene points (with colors and densities) for testing when the model has been trained. By doing so, they successfully reduce the testing time for view synthesis. However, the training time remains the same as the original NeRF. Compared to FastNeRF, our method manages to achieve a shorter time for both training and testing.

Fig. 4: Qualitative comparison with NeRF and NeRF++ on our generated scenes "Bar".

4 Ethics Statement

Our approach to capture and reconstruct the light-field from only a few input images has the immediate utility of many applications, such as augmented, virtual and mixed reality. Our 360° inputs also open up the ability to fully reconstruct and re-render the whole scene at a low cost. Such ability also enables the possibility to reconstruct humans in a scene. The acquisition of such personal information, if without their consent, may lead to privacy and security breaching. Appropriate privacy-preserving steps must be taken to mitigate the potential risk of abusing this technique.

References

1. Adelson, E.H., Bergen, J.R., et al.: The plenoptic function and the elements of early vision, vol. 2. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of ... (1991)
2. Bertel, T., Campbell, N.D., Richardt, C.: Megaparallax: Casual 360° panoramas with motion parallax. IEEE transactions on visualization and computer graphics **25**(5), 1828–1835 (2019)
3. Garbin, S.J., Kowalski, M., Johnson, M., Shotton, J., Valentin, J.: Fastnerf: High-fidelity neural rendering at 200fps. arXiv preprint arXiv:2103.10380 (2021)
4. Huang, J., Chen, Z., Ceylan, D., Jin, H.: 6-dof vr videos with a single 360-camera. In: 2017 IEEE Virtual Reality (VR). pp. 37–44. IEEE (2017)
5. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: European Conference on Computer Vision. pp. 405–421. Springer (2020)
6. Neff, T., Stadlbauer, P., Parger, M., Kurz, A., Chaitanya, C.R.A., Kaplanyan, A., Steinberger, M.: Donerf: Towards real-time rendering of neural radiance fields using depth oracle networks. arXiv preprint arXiv:2103.03231 (2021)
7. Serrano, A., Kim, I., Chen, Z., DiVerdi, S., Gutierrez, D., Hertzmann, A., Masia, B.: Motion parallax for 360 rgbd video. IEEE Transactions on Visualization and Computer Graphics **25**(5), 1817–1827 (2019)
8. Wang, N.H., Solarte, B., Tsai, Y.H., Chiu, W.C., Sun, M.: 360sd-net: 360° stereo depth estimation with learnable cost volume. In: 2020 IEEE International Conference on Robotics and Automation (ICRA). pp. 582–588. IEEE (2020)
9. Zhang, K., Riegler, G., Snavely, N., Koltun, V.: Nerf++: Analyzing and improving neural radiance fields (2020)
10. Zheng, K.C., Kang, S.B., Cohen, M.F., Szeliski, R.: Layered depth panoramas. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8. IEEE (2007)