Global-guided Focal Neural Radiance Field for Large-scale Scene Rendering

Supplement Material

1. Dataset Deails

In our main manuscript, we employ three three street-view datasets: San Francisco Misson Bay, Block_small (MatrixCity), and Block_A (MatrixCity) to evaluate our method performance in street scenes. Here we introduce the details of these three street datasets.

1.1. San Francisco Misson Bay

The San Francisco Misson Bay dataset is a street view dataset proposed by Block-NeRF [3]. This dataset is collected by an autonomous driving collection vehicle equipped with 12 cameras and contains approximately $12k$ images from $1.2km$ length street view. Different cameras in this public dataset are not accurately registered. In order to avoid the error from the unregistered cameras, we opt for the training images of the camera with camera ID of $\mathbf{69}$ among the 12 cameras in the training set. Similarly, we filter out the images with camera ID of $\mathbf{69}$ from the test set for evaluation. It is worth noting that our setting in this dataset is only $1/12$ sparse compared to the original Block-NeRF paper, hence our metrics on the test set are lower than those reported in the original Block-NeRF.

1.2. MatrixCity

MatrixCity is a synthetic large-scale urban dataset built on Unreal Engine [2] where researchers of MatrixCity have constructed two cities in the virtual world: Big-City and Small-City, and have collected $67k$ aerial images and $452k$ street view images, covering a total urban area of $28km^2$. MatrixCity provided a street view benchmark for Block_A and Block_small in small cities, hence we also employ these two datasets for comparison. MatrixCity releases two versions of the two datasets, including the dense version ($0.5m$ sampling interval and the sparse version ($5m$ sampling interval). Since the Blcok_A is a large street scene, a dense sampling with too many images ($20k$ images) would lead to an excessively long training time. Whereas, the Block_small is relatively small, with only $1.5k$ images in the dense version. Thus, we opt for 0.5m interval version for Block_small. All the metrics in our experiments are evaluated on the test views provided by MatrixCity.

2. Detailed Configurations of Experiments

We compare Mega-NeRF [5] and Switch-NeRF [7] on aerial scenes. Since the code for both methods is publicly accessible, we directly utilize the checkpoints provided by these two methods to obtain the comparative results presented in our main document. As the scale of the aerial scenes in our experiment is roughly the same, we use identical configurations for our GF-NeRF across five aerial scenes, with specific settings detailed in Table 1 (GF-NeRF, Aerial-All).

On street-view scenes, we conduct comparisons between F2-NeRF [6], Block-NeRF [3], and GF-NeRF. We employ the official implementation of F2-NeRF, while for Block-NeRF since the code is not publicly available, we implement a version ourselves. Specifically, we implement Block-NeRF in nerfstudio using nerfacto as the base-NeRF [4]. Unlike Mip-NeRF [1], nerfacto employs a hash table to store scene features instead of MLP. As MatrixCity lacks explicit geographical data for block partitioning, in our self-implemented Block-NeRF, we adopt the same block partitioning method as our GF-NeRF, i.e., balanced clustering algorithm. To ensure consistency among blocks during rendering, we calculate the nearest three blocks based on the rendering camera position and render them separately. Finally, we perform weighted compositing based on the reciprocal of the distances to the centers of these three blocks to get the final color.

The geographical extents covered by the three street-view scenes selected in our paper vary significantly. Hence, we use different configurations for each scene, with specific parameters outlined in Table. 1. As the scene Block-A covers the largest area, we utilize a large number of blocks, hash table size, and batch size to ensure that NeRF's capacity is sufficient to store the entire scene and can be adequately trained. Ablation experiments are conducted on the Misson Bay Dataset, with all parameters except the ablation module being consistent with those listed in Table. 1.

References

[1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan.

Table 1. Detailed configurations of compared methods on different scenes.

Method	Dataset	Number of Blocks	Global Steps	Focal (Block) Steps	Batch Size	Hash Table Size (log2)
F2-NeRF	Street-Misson	/	400k	/	8192	21
	Street-Block_A	/	500k	/	16384	23
	Street-Block_small	/	400k	/	8192	21
Block-NeRF	Street-Misson	10	/	30k	8192	21
	Street-Block_A	15	/	30k	16384	23
	Street-Block_small	10	/	30k	8192	21
GF-NeRF	Aerials-All	15	100k	30k	8192	21
	Street-Misson	10	100k	30k	8192	21
	Street-Block_A	15	50k	30k	16384	23
	Street-Block_small	10	100k	30k	8192	21

Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5855–5864, 2021. 1

[2] Yixuan Li, Lihan Jiang, Linning Xu, Yuanbo Xiangli, Zhenzhi Wang, Dahua Lin, and Bo Dai. Matrixcity: A large-scale city dataset for city-scale neural rendering and beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3205–3215, 2023. 1

[3] Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P Srinivasan, Jonathan T Barron, and Henrik Kretzschmar. Block-nerf: Scalable large scene neural view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8248–8258, 2022. 1

[4] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, et al. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 Conference Proceedings*, pages 1–12, 2023. 1

[5] Haithem Turki, Deva Ramanan, and Mahadev Satyanarayanan. Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12922–12931, 2022. 1

[6] Peng Wang, Yuan Liu, Zhaoxi Chen, Lingjie Liu, Ziwei Liu, Taku Komura, Christian Theobalt, and Wenping Wang. F2-nerf: Fast neural radiance field training with free camera trajectories. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4150–4159, 2023. 1

[7] MI Zhenxing and Dan Xu. Switch-nerf: Learning scene decomposition with mixture of experts for large-scale neural radiance fields. In *The Eleventh International Conference on Learning Representations*, 2022. 1