GMT: Enhancing Generalizable Neural Rendering via Geometry-Driven Multi-Reference Texture Transfer –Supplementary Material–

Youngho Yoon*, Hyun-Kurl Jang*, and Kuk-Jin Yoon

Visual Intelligence Lab., KAIST
{dudgh1732, jhg0001, kjyoon}@kaist.ac.kr
https://github.com/yh-yoon/GMT

Abstract. Due to the lack of space in the main paper, we provide more details of the proposed methods and experimental results in the supplementary material. Specifically, in Sec.1, we explain the implementation of extracting alpha values and the training-test step in more detail. In Sec.2, we show ablation analysis according to the various aggregation types. Lastly, Sec.3 presents additional quantitative results of scene-specific novel view synthesis and qualitative comparisons of pixelSplat.

1 More Details of Implementation

1.1 Alpha Values from generalizable NeRFs

Our model uses the alpha value α of the point cloud X^{alpha} obtained from Generalizable NeRF (G-NeRF) models as proxy scene geometry. Among G-NeRF models, the models IBRNet [48], GeoNeRF [23], Neuray [30], MuRF [52] follow the same classical volume rendering process as NeRF, so they estimate the alpha value for each sampled point. On the other hand, GNT [47] calculates point-wise aggregation weights for sampled points in each ray through a ray transformer. This can be interpreted as calculating h_i, or hitting probability, for each point in the equation $c = \prod_{i=1}^{K} h_i c_i$. Therefore, we calculate α_i from h_i through the following equation:

$$\alpha_k = \frac{h_k}{1 - \prod_{i=1}^{k-1} h_i} \tag{1}$$

The calculated α_i can be used as alpha values of our model's alpha point cloud, same as other models.

1.2 Details of Training and Test Step

As mentioned in the implementation details of the main paper, our model uses rendered images and an alpha point cloud of several scenes extracted from a

* Equal contribution.

pre-trained Neuray [30] model as our training dataset. Additionally, in the test step, a single model trained from our training dataset made of Neuray enhances the rendered image extracted by all G-NeRF models(IBRNet [48], GNT [47], GeoNeRF [23], Neuray [30], MuRF [52]). Experimental results performing consistent and significant performance improvement for all G-NeRFs show that our method can robustly learn a multi-reference texture transfer network from given source inputs without being restricted by the G-NeRF model used to generate the training dataset.

2 Ablation Analysis

2.1 Aggregation Types

We present a qualitative result of various feature aggregation methods in Figure 1. Previous studies have commonly employed Multilayer Perceptron (MLP) [30, 48] and Multi-Head Attention (MHA) [47] for feature aggregation. In contrast, our proposed TPFormer introduces a novel approach that aggregates source features while considering the inherent relationship between the source feature and view direction. TPFormer effectively preserves high-quality features from the source images and seamlessly transfers them to the target view. In Figure 1, row (a) illustrates that TPFormer outperforms other aggregation methods in enhancing images, particularly in preserving textures. In row (b) of the figure, we observe that TPFormer effectively removes the blurry artifacts.

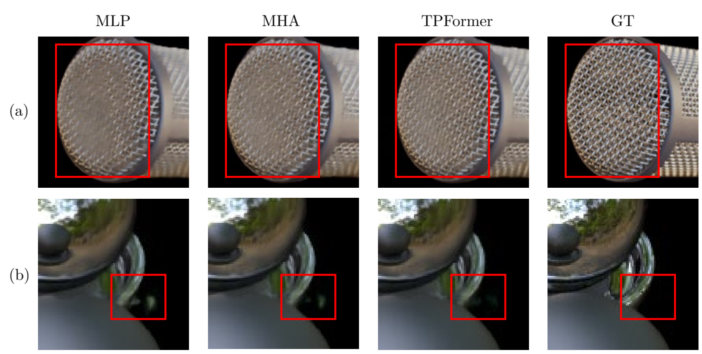

Fig. 1: Qualitative results of different aggregation types. The third column represents the result of the proposed TPFormer.

2.2 The number of reference images

We examine the impact of the number of reference images to verify whether reference images are appropriately used. As shown in Fig. 2, performance gradually improves as the number of reference images increases. This result indicates

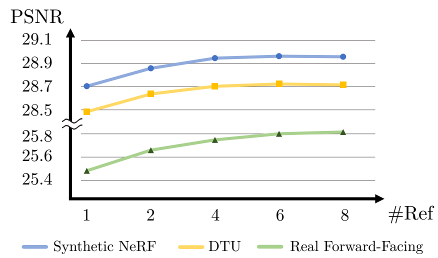

Fig. 2: Comparison of PSNR according to the number of reference images.

Table 1: Results on per-scene optimization NeRFs.

Model	DVGO [43]			TensoRF [5]		
	PSNR(↑)	SSIM(↑)	LPIPS(↓)	PSNR(↑)	SSIM(↑)	LPIPS(↓)
w/o Ours	26.56	0.847	0.165	25.97	0.830	0.177
Ours	**26.82**	**0.863**	**0.144**	**26.23**	**0.848**	**0.153**

that our model effectively chooses relevant features from multiple images and incorporates them with precision.

3 Additional Experimental Results

3.1 Additional Results with per-scene optimized NeRFs

We conducted experiments with DVGO [43] and TensoRF [5] to verify the ability of our model to enhance the results of per-scene optimized NeRFs. As in Table 1, our model shows improved performance on the Real Forward-Facing dataset.

3.2 Additional Quantitative Results

We show quantitative results of scene-specific novel view synthesis of generalizable NeRFs (IBRNet [48], GNT [47], GeoNeRF [23], Neuray [30], MuRF [52]) and our models for three datasets. The results presented in Table 2 correspond to DTU, Table 3 reflects the results for Synthetic NeRF, and Table 4 showcases the results for the Real Forward-Facing dataset. We also show scene-specific novel-view synthesis results of four image enhancement models (C2-Matching [22], MRefSR [61], NeRFLiX [66], Ours) on the three datasets. The results presented in Table 5 correspond to DTU, Table 6 reflects the results for Synthetic NeRF, and Table 7 showcases the results for the Real Forward-Facing dataset.

3.3 Additional Qualitative Comparisons of pixelSplat

We additionally present a qualitative comparison of the novel view synthesis produced by pixelSplat [3], both with and without ours. We show the result of RealEstate10k and ACID datasets in Fig. 3 and Fig. 4. The results reveal that the utilization of ours consistently yields superior texture quality and fewer instances of blurry artifacts compared to scenarios where ours is not employed.

References

1. Cao, A., Rockwell, C., Johnson, J.: Fwd: Real-time novel view synthesis with forward warping and depth. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15713–15724 (2022)
2. Cao, J., Liang, J., Zhang, K., Li, Y., Zhang, Y., Wang, W., Gool, L.V.: Reference-based image super-resolution with deformable attention transformer. In: European conference on computer vision. pp. 325–342. Springer (2022)
3. Charatan, D., Li, S., Tagliasacchi, A., Sitzmann, V.: pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. arXiv preprint arXiv:2312.12337 (2023)
4. Chaurasia, G., Duchene, S., Sorkine-Hornung, O., Drettakis, G.: Depth synthesis and local warps for plausible image-based navigation. ACM Transactions on Graphics (TOG) **32**(3), 1–12 (2013)
5. Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H.: Tensorf: Tensorial radiance fields. In: European Conference on Computer Vision. pp. 333–350. Springer (2022)
6. Chen, A., Xu, Z., Zhao, F., Zhang, X., Xiang, F., Yu, J., Su, H.: Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14124–14133 (2021)
7. Choi, I., Gallo, O., Troccoli, A., Kim, M.H., Kautz, J.: Extreme view synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7781–7790 (2019)
8. Choy, C., Gwak, J., Savarese, S.: 4d spatio-temporal convnets: Minkowski convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3075–3084 (2019)
9. Debevec, P., Yu, Y., Borshukov, G.: Efficient view-dependent image-based rendering with projective texture-mapping. In: Rendering Techniques' 98: Proceedings of the Eurographics Workshop in Vienna, Austria, June 29—July 1, 1998 9. pp. 105–116. Springer (1998)
10. Deng, K., Liu, A., Zhu, J.Y., Ramanan, D.: Depth-supervised nerf: Fewer views and faster training for free. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12882–12891 (2022)
11. Downs, L., Francis, A., Koenig, N., Kinman, B., Hickman, R., Reymann, K., McHugh, T.B., Vanhoucke, V.: Google scanned objects: A high-quality dataset of 3d scanned household items. In: 2022 International Conference on Robotics and Automation (ICRA). pp. 2553–2560. IEEE (2022)
12. Du, Y., Smith, C., Tewari, A., Sitzmann, V.: Learning to render novel views from wide-baseline stereo pairs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4970–4980 (2023)
13. Flynn, J., Broxton, M., Debevec, P., DuVall, M., Fyffe, G., Overbeck, R., Snavely, N., Tucker, R.: Deepview: View synthesis with learned gradient descent. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2367–2376 (2019)

14. Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B., Kanazawa, A.: Plenoxels: Radiance fields without neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5501–5510 (2022)
15. Gatys, L., Ecker, A.S., Bethge, M.: Texture synthesis using convolutional neural networks. Advances in neural information processing systems **28** (2015)
16. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2414–2423 (2016)
17. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The Lumigraph. Association for Computing Machinery, New York, NY, USA, 1 edn. (2023), `https://doi.org/10.1145/3596711.3596760`
18. Hedman, P., Philip, J., Price, T., Frahm, J.M., Drettakis, G., Brostow, G.: Deep blending for free-viewpoint image-based rendering. ACM Transactions on Graphics (ToG) **37**(6), 1–15 (2018)
19. Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th international conference on pattern recognition. pp. 2366–2369. IEEE (2010)
20. Huang, J., Thies, J., Dai, A., Kundu, A., Jiang, C., Guibas, L.J., Nießner, M., Funkhouser, T., et al.: Adversarial texture optimization from rgb-d scans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1559–1568 (2020)
21. Jensen, R., Dahl, A., Vogiatzis, G., Tola, E., Aanæs, H.: Large scale multi-view stereopsis evaluation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 406–413 (2014)
22. Jiang, Y., Chan, K.C., Wang, X., Loy, C.C., Liu, Z.: Robust reference-based super-resolution via c2-matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2103–2112 (2021)
23. Johari, M.M., Lepoittevin, Y., Fleuret, F.: Geonerf: Generalizing nerf with geometry priors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18365–18375 (2022)
24. Kalantari, N.K., Wang, T.C., Ramamoorthi, R.: Learning-based view synthesis for light field cameras. ACM Transactions on Graphics (TOG) **35**(6), 1–10 (2016)
25. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4) (2023)
26. Levoy, M., Hanrahan, P.: Light field rendering. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques. p. 31–42. SIGGRAPH '96, Association for Computing Machinery, New York, NY, USA (1996). `https://doi.org/10.1145/237170.237199`, `https://doi.org/10.1145/237170.237199`
27. Li, Y., Luo, Y., Lu, J.: Reference-guided deep deblurring via a selective attention network. Applied Intelligence pp. 1–13 (2022)
28. Liu, A., Tucker, R., Jampani, V., Makadia, A., Snavely, N., Kanazawa, A.: Infinite nature: Perpetual view generation of natural scenes from a single image. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14458–14467 (2021)
29. Liu, C., Hua, Z., Li, J.: Reference-based dual-task framework for motion deblurring. The Visual Computer pp. 1–15 (2023)
30. Liu, Y., Peng, S., Liu, L., Wang, Q., Wang, P., Theobalt, C., Zhou, X., Wang, W.: Neural rays for occlusion-aware image-based rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7824–7833 (2022)

31. Lu, L., Li, W., Tao, X., Lu, J., Jia, J.: Masa-sr: Matching acceleration and spatial adaptation for reference-based image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6368–6377 (2021)
32. Mildenhall, B., Srinivasan, P.P., Ortiz-Cayon, R., Kalantari, N.K., Ramamoorthi, R., Ng, R., Kar, A.: Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. ACM Transactions on Graphics (TOG) **38**(4), 1–14 (2019)
33. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
34. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. arXiv preprint arXiv:2201.05989 (2022)
35. Penner, E., Zhang, L.: Soft 3d reconstruction for view synthesis. ACM Transactions on Graphics (TOG) **36**(6), 1–11 (2017)
36. Pesavento, M., Volino, M., Hilton, A.: Attention-based multi-reference learning for image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14697–14706 (2021)
37. Riegler, G., Koltun, V.: Free view synthesis. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX 16. pp. 623–640. Springer (2020)
38. Roessle, B., Barron, J.T., Mildenhall, B., Srinivasan, P.P., Nießner, M.: Dense depth priors for neural radiance fields from sparse input views. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12892–12901 (2022)
39. Shum, H., Kang, S.B.: Review of image-based rendering techniques. In: Visual Communications and Image Processing 2000. vol. 4067, pp. 2–13. SPIE (2000)
40. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
41. Suhail, M., Esteves, C., Sigal, L., Makadia, A.: Generalizable patch-based neural rendering. In: European Conference on Computer Vision. pp. 156–174. Springer (2022)
42. Suhail, M., Esteves, C., Sigal, L., Makadia, A.: Light field neural rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8269–8279 (2022)
43. Sun, C., Sun, M., Chen, H.T.: Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5459–5469 (2022)
44. Thies, J., Zollhöfer, M., Nießner, M.: Deferred neural rendering: Image synthesis using neural textures. Acm Transactions on Graphics (TOG) **38**(4), 1–12 (2019)
45. Thies, J., Zollhöfer, M., Theobalt, C., Stamminger, M., Nießner, M.: Ignor: Image-guided neural object rendering. arXiv preprint arXiv:1811.10720 (2018)
46. Trevithick, A., Yang, B.: Grf: Learning a general radiance field for 3d representation and rendering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 15182–15192 (2021)
47. Wang, P., Chen, X., Chen, T., Venugopalan, S., Wang, Z., et al.: Is attention all nerf needs? arXiv preprint arXiv:2207.13298 (2022)
48. Wang, Q., Wang, Z., Genova, K., Srinivasan, P.P., Zhou, H., Barron, J.T., Martin-Brualla, R., Snavely, N., Funkhouser, T.: Ibrnet: Learning multi-view image-based rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4690–4699 (2021)

49. Wei, Y., Liu, S., Rao, Y., Zhao, W., Lu, J., Zhou, J.: Nerfingmvs: Guided optimization of neural radiance fields for indoor multi-view stereo. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5610–5619 (2021)
50. Wizadwongsa, S., Phongthawee, P., Yenphraphai, J., Suwajanakorn, S.: Nex: Real-time view synthesis with neural basis expansion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8534–8543 (2021)
51. Xia, B., Tian, Y., Hang, Y., Yang, W., Liao, Q., Zhou, J.: Coarse-to-fine embedded patchmatch and multi-scale dynamic aggregation for reference-based super-resolution. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 2768–2776 (2022)
52. Xu, H., Chen, A., Chen, Y., Sakaridis, C., Zhang, Y., Pollefeys, M., Geiger, A., Yu, F.: Murf: Multi-baseline radiance fields. arXiv preprint arXiv:2312.04565 (2023)
53. Xu, Q., Xu, Z., Philip, J., Bi, S., Shu, Z., Sunkavalli, K., Neumann, U.: Point-nerf: Point-based neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5438–5448 (2022)
54. Xu, Z., Bi, S., Sunkavalli, K., Hadap, S., Su, H., Ramamoorthi, R.: Deep view synthesis from sparse photometric images. ACM Transactions on Graphics (ToG) **38**(4), 1–13 (2019)
55. Yang, F., Yang, H., Fu, J., Lu, H., Guo, B.: Learning texture transformer network for image super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5791–5800 (2020)
56. Yang, H., Hong, L., Li, A., Hu, T., Li, Z., Lee, G.H., Wang, L.: Contranerf: Generalizable neural radiance fields for synthetic-to-real novel view synthesis via contrastive learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16508–16517 (2023)
57. Yao, Y., Luo, Z., Li, S., Fang, T., Quan, L.: Mvsnet: Depth inference for unstructured multi-view stereo. In: Proceedings of the European conference on computer vision (ECCV). pp. 767–783 (2018)
58. Youngho, Y., Kuk-Jin, Y.: Cross-guided optimization of radiance fields with multi-view image super-resolution for high-resolution novel view synthesis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. p. 12428–12438 (2023)
59. Yu, A., Li, R., Tancik, M., Li, H., Ng, R., Kanazawa, A.: Plenoctrees for real-time rendering of neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5752–5761 (2021)
60. Yu, A., Ye, V., Tancik, M., Kanazawa, A.: pixelnerf: Neural radiance fields from one or few images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4578–4587 (2021)
61. Zhang, L., Li, X., He, D., Li, F., Ding, E., Zhang, Z.: Lmr: A large-scale multi-reference dataset for reference-based super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13118–13127 (2023)
62. Zhang, L., Li, X., He, D., Li, F., Wang, Y., Zhang, Z.: Rrsr: Reciprocal reference-based image super-resolution with progressive feature alignment and selection. In: European Conference on Computer Vision. pp. 648–664. Springer (2022)
63. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
64. Zhang, Z., Wang, Z., Lin, Z., Qi, H.: Image super-resolution by neural texture transfer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7982–7991 (2019)

65. Zheng, H., Ji, M., Wang, H., Liu, Y., Fang, L.: Crossnet: An end-to-end reference-based super resolution network using cross-scale warping. In: Proceedings of the European conference on computer vision (ECCV). pp. 88–104 (2018)
66. Zhou, K., Li, W., Wang, Y., Hu, T., Jiang, N., Han, X., Lu, J.: Nerflix: High-quality neural view synthesis by learning a degradation-driven inter-viewpoint mixer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12363–12374 (2023)
67. Zhou, T., Tucker, R., Flynn, J., Fyffe, G., Snavely, N.: Stereo magnification: Learning view synthesis using multiplane images. arXiv preprint arXiv:1805.09817 (2018)
68. Zou, H., Suganuma, M., Okatani, T.: Reference-based motion blur removal: Learning to utilize sharpness in the reference image. arXiv preprint arXiv:2307.02875 (2023)

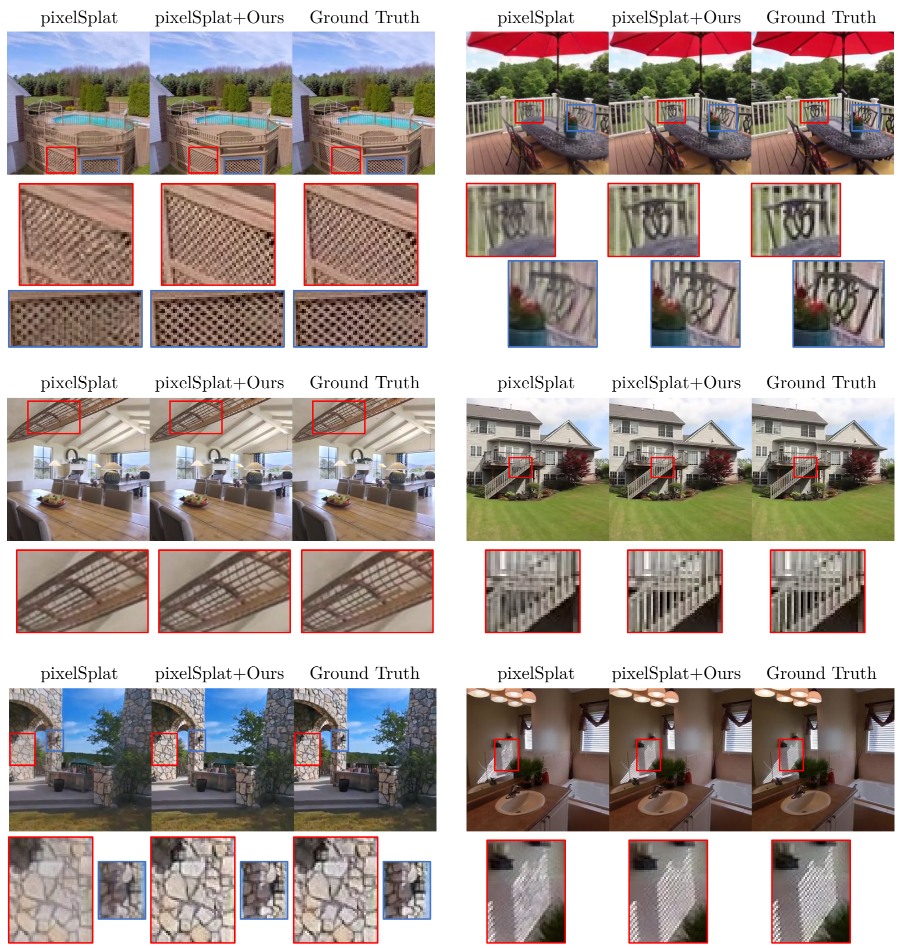

Fig. 3: Qualitative comparisons of pixelSplat model with and without incorporating our proposed model across on RealEstate10k dataset.

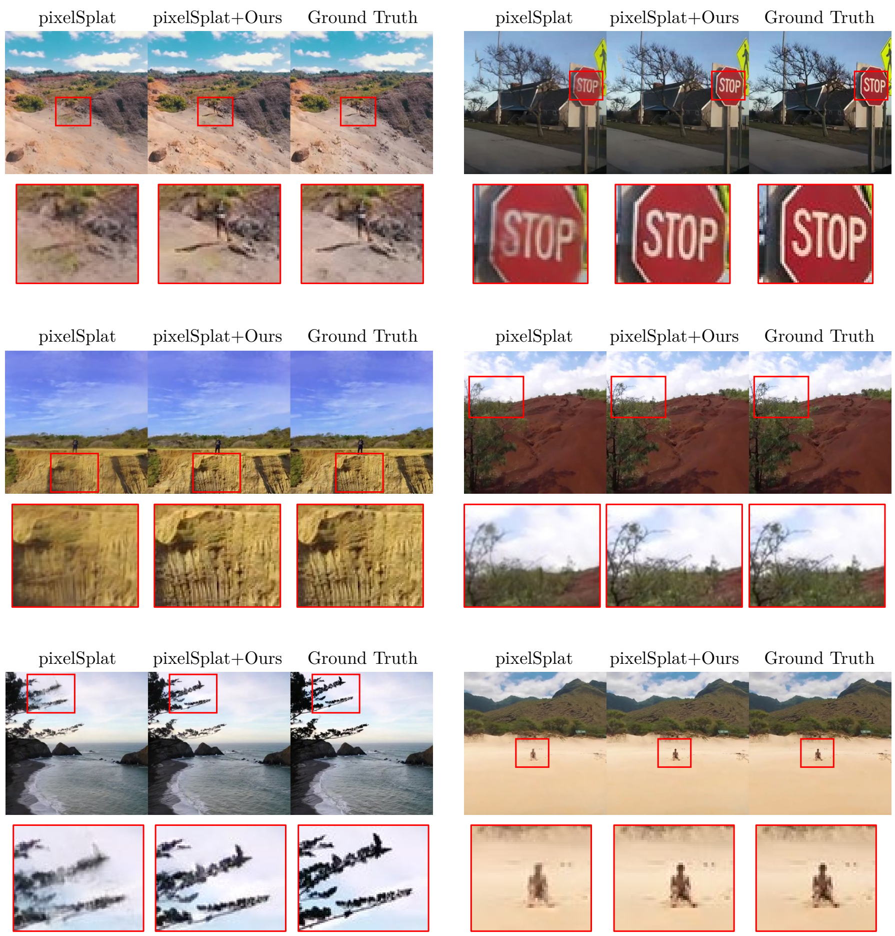

Fig. 4: Qualitative comparisons of pixelSplat model with and without incorporating our proposed model across on ACID dataset.

Table 2: Per-scene quantitative comparison on DTU dataset. Ours consistently exhibits significant improvements across scenes.

Method	Birds	Bricks	Snowman	Tools	Avg.
PSNR(↑)					
IBRNet [48]	29.66	25.50	28.64	22.27	26.76
GNT [47]	27.89	23.81	28.21	21.12	25.46
GeoNeRF [23]	-				-
Neuray [30]	32.78	26.95	29.12	23.16	28.37
MuRF [52]	27.18	24.27	27.07	20.20	24.87
Ours+IBRNet [48]	29.82	25.92	28.72	22.42	26.96
	(0.16 ↑)	(0.42 ↑)	(0.08↑)	(0.15 ↑)	(0.20 ↑)
Ours+GNT [47]	28.07	24.05	28.24	21.23	25.60
	(0.18↑)	(0.24↑)	(0.03 ↑)	(0.11 ↑)	(0.14↑)
Ours+GeoNeRF [23]	-				-
Ours+Neuray [30]	33.44	27.35	29.19	23.31	28.72
	(0.66 ↑)	(0.40 ↑)	(0.07↑)	(0.15 ↑)	(0.35 ↑)
Ours+MuRF [52]	27.28	24.32	27.05	20.21	24.91
	(0.10 ↑)	(0.05 ↑)	(0.02↓)	(0.01 ↑)	(0.04 ↑)
SSIM(↑)					
IBRNet [48]	0.922	0.821	0.917	0.843	0.879
GNT [47]	0.816	0.797	0.890	0.769	0.818
GeoNeRF [23]	-				-
Neuray [30]	0.943	0.869	0.926	0.872	0.906
MuRF [52]	0.884	0.889	0.907	0.794	0.870
Ours+IBRNet [48]	0.931	0.847	0.922	0.859	0.893
	(0.009↑)	(0.026 ↑)	(0.005 ↑)	(0.016 ↑)	(0.014 ↑)
Ours+GNT [47]	0.828	0.823	0.895	0.784	0.832
	(0.012↑)	(0.026 ↑)	(0.005↑)	(0.015 ↑)	(0.014 ↑)
Ours+GeoNeRF [23]	-				-
Ours+Neuray [30]	0.952	0.897	0.932	0.888	0.920
	(0.009↑)	(0.028 ↑)	(0.006↑)	(0.016 ↑)	(0.014 ↑)
Ours+MuRF [52]	0.885	0.894	0.906	0.798	0.872
	(0.001↑)	(0.005 ↑)	(0.001↓)	(0.004 ↑)	(0.002 ↑)
LPIPS(↓)					
IBRNet [48]	0.113	0.188	0.120	0.131	0.136
GNT [47]	0.174	0.219	0.141	0.150	0.171
GeoNeRF [23]	-				-
Neuray [30]	0.088	0.154	0.104	0.108	0.112
MuRF [52]	0.172	0.195	0.157	0.212	0.183
Ours+IBRNet [48]	0.095	0.173	0.112	0.125	0.124
	(0.018 ↓)	(0.015 ↓)	(0.008↓)	(0.006 ↓)	(0.012 ↓)
Ours+GNT [47]	0.147	0.201	0.132	0.144	0.155
	(0.027 ↓)	(0.018 ↓)	(0.009↓)	(0.006 ↓)	(0.016 ↓)
Ours+GeoNeRF [23]	-				-
Ours+Neuray [30]	0.076	0.136	0.097	0.101	0.101
	(0.012 ↓)	(0.018 ↓)	(0.007 ↓)	(0.007 ↓)	(0.011 ↓)
Ours+MuRF [52]	0.171	0.189	0.164	0.199	0.180
	(0.001 ↓)	(0.006 ↓)	(0.007 ↑)	(0.013 ↓)	(0.003 ↓)

Table 3: Per-scene quantitative comparison on synthetic NeRF dataset. Ours consistently exhibits significant improvements across most scenes.

Method	Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship	Avg.
				PSNR(↑)					
IBRNet [48]	24.15	20.80	21.59	31.64	24.86	22.18	28.26	26.79	25.03
GNT [47]	22.14	19.48	24.33	26.78	21.54	20.71	26.97	22.97	23.11
GeoNeRF [23]	30.16	24.57	26.13	33.90	30.39	28.91	33.72	28.91	29.59
Neuray [30]	29.92	23.37	24.77	34.07	28.58	26.78	31.49	27.93	28.36
MuRF [52]	21.81	18.59	20.26	26.68	20.65	19.49	26.44	24.17	22.26
Ours+IBRNet [48]	24.38	20.88	21.54	32.30	25.16	22.20	28.55	27.28	25.29
	(0.23 ↑)	(0.08 ↑)	(0.05↓)	(0.66 ↑)	(0.30 ↑)	(0.02 ↑)	(0.29 ↑)	(0.49 ↑)	(0.26 ↑)
Ours+GNT [47]	22.23	19.53	24.53	26.97	21.65	20.63	27.22	23.06	23.23
	(0.09 ↑)	(0.05 ↑)	(0.20↑)	(0.19 ↑)	(0.11 ↑)	(0.08 ↓)	(0.25 ↑)	(0.09 ↑)	(0.12 ↑)
Ours+GeoNeRF [23]	30.59	24.57	26.09	34.66	30.66	28.85	33.99	29.08	29.81
	(0.43 ↑)	(0.00 ↑)	(0.04↓)	(0.76 ↑)	(0.27 ↑)	(0.06 ↓)	(0.27 ↑)	(0.17 ↑)	(0.22 ↑)
Ours+Neuray [30]	31.15	23.62	25.13	34.84	29.31	26.93	32.39	28.30	28.96
	(1.23↑)	(0.24 ↑)	(0.36 ↑)	(0.77 ↑)	(0.73 ↑)	(0.15 ↑)	(0.90 ↑)	(0.37 ↑)	(0.60 ↑)
Ours+MuRF [52]	21.96	18.65	20.28	26.83	20.73	19.52	26.61	24.32	22.36
	(0.15↑)	(0.06 ↑)	(0.02 ↑)	(0.15 ↑)	(0.08 ↑)	(0.03 ↑)	(0.17 ↑)	(0.15 ↑)	(0.10 ↑)
				SSIM(↑)					
IBRNet [48]	0.933	0.867	0.882	0.964	0.905	0.871	0.948	0.831	0.900
GNT [47]	0.752	0.704	0.802	0.821	0.740	0.720	0.849	0.720	0.763
GeoNeRF [23]	0.961	0.913	0.921	0.959	0.949	0.923	0.978	0.858	0.933
Neuray [30]	0.964	0.909	0.908	0.972	0.940	0.915	0.970	0.848	0.928
MuRF [52]	0.600	0.570	0.494	0.694	0.626	0.572	0.645	0.693	0.612
Ours+IBRNet [48]	0.942	0.881	0.888	0.968	0.915	0.879	0.955	0.844	0.909
	(0.009 ↑)	(0.014 ↑)	(0.006 ↑)	(0.004 ↑)	(0.010 ↑)	(0.008 ↑)	(0.007 ↑)	(0.013 ↑)	(0.009 ↑)
Ours+GNT [47]	0.764	0.720	0.810	0.826	0.752	0.728	0.856	0.733	0.773
	(0.012 ↑)	(0.016 ↑)	(0.008 ↑)	(0.005 ↑)	(0.012 ↑)	(0.008 ↑)	(0.007 ↑)	(0.013 ↑)	(0.010 ↑)
Ours+GeoNeRF [23]	0.968	0.926	0.929	0.969	0.955	0.937	0.982	0.871	0.942
	(0.007 ↑)	(0.013 ↑)	(0.008↑)	(0.010 ↑)	(0.006 ↑)	(0.014 ↑)	(0.004 ↑)	(0.013 ↑)	(0.009 ↑)
Ours+Neuray [30]	0.972	0.920	0.919	0.974	0.948	0.924	0.976	0.858	0.936
	(0.008 ↑)	(0.011↑)	(0.011 ↑)	(0.002 ↑)	(0.008 ↑)	(0.009 ↑)	(0.006 ↑)	(0.010 ↑)	(0.008↑)
Ours+MuRF [52]	0.604	0.572	0.495	0.696	0.630	0.574	0.647	0.697	0.614
	(0.004 ↑)	(0.002↑)	(0.001 ↑)	(0.002 ↑)	(0.004 ↑)	(0.002 ↑)	(0.002 ↑)	(0.004 ↑)	(0.002↑)
				LPIPS(↓)					
IBRNet [48]	0.064	0.121	0.115	0.052	0.099	0.127	0.054	0.183	0.102
GNT [47]	0.108	0.171	0.101	0.100	0.152	0.170	0.076	0.249	0.141
GeoNeRF [23]	0.043	0.079	0.081	0.053	0.057	0.084	0.025	0.137	0.071
Neuray [30]	0.036	0.083	0.088	0.038	0.063	0.083	0.025	0.154	0.071
MuRF [52]	0.191	0.239	0.231	0.176	0.243	0.260	0.141	0.316	0.225
Ours+IBRNet [48]	0.052	0.104	0.103	0.042	0.083	0.114	0.042	0.175	0.089
	(0.012 ↓)	(0.017 ↓)	(0.012 ↓)	(0.010 ↓)	(0.016 ↓)	(0.013 ↓)	(0.012 ↓)	(0.008 ↓)	(0.013 ↓)
Ours+GNT [47]	0.088	0.141	0.086	0.088	0.129	0.157	0.064	0.238	0.124
	(0.020 ↓)	(0.030 ↓)	(0.015 ↓)	(0.012 ↓)	(0.023 ↓)	(0.013 ↓)	(0.012 ↓)	(0.011 ↓)	(0.017 ↓)
Ours+GeoNeRF [23]	0.028	0.058	0.058	0.031	0.041	0.058	0.015	0.131	0.052
	(0.015 ↓)	(0.021 ↓)	(0.023↓)	(0.022 ↓)	(0.016 ↓)	(0.026 ↓)	(0.010 ↓)	(0.006 ↓)	(0.019 ↓)
Ours+Neuray [30]	0.028	0.070	0.069	0.033	0.050	0.073	0.020	0.149	0.062
	(0.008 ↓)	(0.013 ↓)	(0.019↓)	(0.005 ↓)	(0.013 ↓)	(0.010 ↓)	(0.005 ↓)	(0.005 ↓)	(0.010 ↓)
Ours+MuRF [52]	0.182	0.229	0.229	0.171	0.233	0.255	0.130	0.311	0.217
	(0.009 ↓)	(0.010 ↓)	(0.002↓)	(0.005 ↓)	(0.010 ↓)	(0.005 ↓)	(0.011 ↓)	(0.005 ↓)	(0.008 ↓)

Table 4: Per-scene quantitative comparison on Real Forward-Facing dataset. Ours consistently exhibits significant improvements across most scenes.

Method	Fern	Flower	Fortress	Horns	Leaves	Orchids	Room	Trex	Avg.
PSNR(↑)									
IBRNet [48]	23.45	26.29	29.89	25.82	19.79	19.01	28.99	23.76	25.19
GNT [47]	24.14	25.77	30.51	26.36	19.81	18.52	29.66	24.55	25.54
GeoNeRF [23]	24.10	27.89	30.16	26.59	20.28	20.09	28.50	23.52	25.64
Neuray [30]	23.54	27.08	29.02	26.70	19.58	19.33	29.18	24.18	25.43
MuRF [52]	23.40	28.89	30.28	27.16	21.58	21.61	30.07	24.22	26.43
Ours+IBRNet [48]	23.82	26.38	30.23	26.37	19.91	19.11	29.72	24.20	25.57
	(0.37 ↑)	(0.09 ↑)	(0.34 ↑)	(0.55 ↑)	(0.12 ↑)	(0.10 ↑)	(0.73 ↑)	(0.44 ↑)	(0.38 ↑)
Ours+GNT [47]	24.45	25.88	30.73	26.61	19.95	18.56	30.16	25.01	25.81
	(0.31 ↑)	(0.11 ↑)	(0.22 ↑)	(0.25 ↑)	(0.14 ↑)	(0.04 ↑)	(0.50 ↑)	(0.46 ↑)	(0.27 ↑)
Ours+GeoNeRF [23]	24.15	27.91	30.17	26.77	20.31	20.06	29.07	23.78	25.80
	(0.05 ↑)	(0.02 ↑)	(0.01 ↑)	(0.18 ↑)	(0.03 ↑)	(0.03 ↓)	(0.57 ↑)	(0.26 ↑)	(0.16 ↑)
Ours+Neuray [30]	23.90	27.17	29.33	27.32	19.64	19.46	29.84	24.66	25.82
	(0.36 ↑)	(0.09 ↑)	(0.31 ↑)	(0.62 ↑)	(0.06 ↑)	(0.13 ↑)	(0.66 ↑)	(0.48 ↑)	(0.39↑)
Ours+MuRF [52]	23.55	29.03	30.62	27.64	21.66	21.58	31.19	24.58	26.81
	(0.15 ↑)	(0.14 ↑)	(0.34 ↑)	(0.48 ↑)	(0.08 ↑)	(0.03 ↓)	(1.12 ↑)	(0.36 ↑)	(0.38↑)
SSIM(↑)									
IBRNet [48]	0.761	0.847	0.878	0.856	0.696	0.610	0.934	0.840	0.822
GNT [47]	0.791	0.838	0.891	0.876	0.696	0.623	0.942	0.868	0.835
GeoNeRF [23]	0.796	0.873	0.889	0.892	0.713	0.655	0.939	0.874	0.847
Neuray [30]	0.768	0.858	0.884	0.878	0.682	0.629	0.941	0.855	0.833
MuRF [52]	0.828	0.939	0.946	0.926	0.840	0.820	0.968	0.903	0.908
Ours+IBRNet [48]	0.788	0.852	0.891	0.879	0.715	0.627	0.944	0.863	0.839
	(0.027 ↑)	(0.005 ↑)	(0.013 ↑)	(0.023 ↑)	(0.019 ↑)	(0.017 ↑)	(0.010 ↑)	(0.023↑)	(0.017 ↑)
Ours+GNT [47]	0.814	0.845	0.903	0.891	0.721	0.638	0.949	0.888	0.850
	(0.023 ↑)	(0.007 ↑)	(0.012 ↑)	(0.015 ↑)	(0.025 ↑)	(0.015 ↑)	(0.007 ↑)	(0.020 ↑)	(0.015 ↑)
Ours+GeoNeRF [23]	0.808	0.876	0.893	0.899	0.728	0.663	0.949	0.890	0.857
	(0.012 ↑)	(0.003 ↑)	(0.004 ↑)	(0.007 ↑)	(0.015 ↑)	(0.008 ↑)	(0.010 ↑)	(0.016 ↑)	(0.010 ↑)
Ours+Neuray [30]	0.797	0.863	0.897	0.901	0.702	0.648	0.950	0.877	0.850
	(0.029 ↑)	(0.005 ↑)	(0.013 ↑)	(0.023 ↑)	(0.020 ↑)	(0.019 ↑)	(0.009 ↑)	(0.022 ↑)	(0.017 ↑)
Ours+MuRF [52]	0.839	0.940	0.951	0.936	0.850	0.822	0.975	0.915	0.915
	(0.011 ↑)	(0.001 ↑)	(0.005 ↑)	(0.010 ↑)	(0.010 ↑)	(0.002 ↑)	(0.007 ↑)	(0.012 ↑)	(0.007 ↑)
LPIPS(↓)									
IBRNet [48]	0.217	0.152	0.127	0.162	0.228	0.291	0.106	0.177	0.173
GNT [47]	0.207	0.174	0.148	0.162	0.239	0.283	0.109	0.171	0.177
GeoNeRF [23]	0.183	0.131	0.108	0.124	0.218	0.246	0.101	0.163	0.150
Neuray [30]	0.209	0.134	0.126	0.145	0.235	0.265	0.093	0.167	0.161
MuRF [52]	0.202	0.118	0.123	0.145	0.164	0.163	0.095	0.154	0.141
Ours+IBRNet [48]	0.192	0.145	0.115	0.140	0.205	0.279	0.088	0.150	0.154
	(0.025 ↓)	(0.007 ↓)	(0.012 ↓)	(0.022 ↓)	(0.023 ↓)	(0.012 ↓)	(0.018 ↓)	(0.027 ↓)	(0.019 ↓)
Ours+GNT [47]	0.183	0.163	0.129	0.139	0.218	0.273	0.091	0.143	0.157
	(0.024 ↓)	(0.011 ↓)	(0.019 ↓)	(0.023 ↓)	(0.021 ↓)	(0.010 ↓)	(0.018 ↓)	(0.028↓)	(0.020 ↓)
Ours+GeoNeRF [23]	0.169	0.126	0.108	0.114	0.195	0.236	0.081	0.140	0.137
	(0.014 ↓)	(0.005 ↓)	(0.000 ↓)	(0.010 ↓)	(0.023 ↓)	(0.010 ↓)	(0.020 ↓)	(0.023 ↓)	(0.013 ↓)
Ours+Neuray [30]	0.184	0.129	0.113	0.120	0.208	0.248	0.079	0.141	0.142
	(0.025↓)	(0.005 ↓)	(0.013 ↓)	(0.025 ↓)	(0.027 ↓)	(0.017 ↓)	(0.014 ↓)	(0.026 ↓)	(0.019 ↓)
Ours+MuRF [52]	0.198	0.127	0.119	0.141	0.154	0.165	0.093	0.153	0.139
	(0.004↓)	(0.009 ↑)	(0.004 ↓)	(0.004 ↓)	(0.010 ↓)	(0.002 ↑)	(0.002 ↓)	(0.001 ↓)	(0.002 ↓)

Table 5: Per-scene quantitative comparison of our model with other reference-based image enhancement models on the DTU dataset. We use Neuray as the G-NeRF baseline for all models. **Bold** indicates the best results, and underline indicates the second best results.

Method	Birds	Bricks	Snowman	Tools	Avg.
PSNR(↑)					
C2-Matching [22]	32.80	27.01	29.08	23.13	28.38
MRefSR [61]	33.12	27.13	29.11	23.18	28.52
NeRFLiX [66]	33.10	27.22	29.08	23.24	28.54
Ours	**33.44**	**27.35**	**29.19**	**23.31**	**28.72**
SSIM(↑)					
C2-Matching [22]	0.945	0.878	0.926	0.877	0.910
MRefSR [61]	0.937	0.878	0.919	0.869	0.903
NeRFLiX [66]	0.939	0.882	0.921	0.873	0.906
Ours	**0.952**	**0.897**	**0.932**	**0.888**	**0.920**
LPIPS(↓)					
C2-Matching [22]	0.082	0.146	0.100	**0.101**	0.105
MRefSR [61]	0.078	0.143	0.100	0.106	0.104
NeRFLiX [66]	0.077	0.140	0.098	0.105	0.103
Ours	**0.076**	**0.136**	**0.097**	**0.101**	**0.101**

Table 6: Per-scene quantitative comparison of our model with other reference-based enhancement models on the synthetic NeRF dataset. We use Neuray as the G-NeRF baseline for all models. **Bold** indicates the best results, and underline indicates the second best results.

Method	Chair	Drums	Ficus	Hotdog	Lego	Materials	Mic	Ship	Avg.
PSNR(↑)									
C2-Matching [22]	30.22	23.41	24.75	34.23	28.69	26.85	31.74	28.01	28.49
MRefSR [61]	30.54	23.45	24.92	34.50	28.89	26.89	31.81	28.10	28.64
NeRFLiX [66]	30.82	23.53	25.07	34.64	29.01	**26.97**	32.12	28.14	28.79
Ours	**31.15**	**23.62**	**25.13**	**34.84**	**29.31**	26.93	**32.39**	**28.30**	**28.96**
SSIM(↑)									
C2-Matching [22]	0.963	0.900	0.910	0.970	0.942	0.916	0.971	0.849	0.928
MRefSR [61]	0.951	0.901	0.897	0.958	0.930	0.907	0.952	0.844	0.917
NeRFLiX [66]	0.963	0.907	0.908	0.966	0.938	0.913	0.964	0.847	0.926
Ours	**0.972**	**0.920**	**0.919**	**0.974**	**0.948**	**0.924**	**0.976**	**0.858**	**0.936**
LPIPS(↓)									
C2-Matching [22]	0.033	0.078	0.082	0.036	0.058	0.079	0.024	0.151	0.068
MRefSR [61]	0.039	0.080	0.082	0.041	0.060	0.080	0.033	0.153	0.071
NeRFLiX [66]	**0.028**	0.073	**0.069**	**0.032**	0.051	**0.071**	0.021	**0.148**	**0.062**
Ours	**0.028**	**0.070**	**0.069**	0.033	**0.050**	0.073	**0.020**	0.149	**0.062**

Table 7: Per-scene quantitative comparison of our model with other reference-based enhancement models on the Real Forward-Facing dataset. We use Neuray as the G-NeRF baseline for all models. **Bold** indicates the best results, and underline indicates the second best results.

Method	Fern	Flower	Fortress	Horns	Leaves	Orchids	Room	Trex	Avg.
PSNR(↑)									
C2-Matching [22]	23.62	27.08	29.13	26.89	19.57	19.32	29.47	24.30	25.55
MRefSR [61]	23.68	27.10	29.22	27.10	19.56	19.36	29.63	24.36	25.64
NeRFLiX [66]	23.70	27.10	29.13	27.11	19.59	19.36	29.67	24.41	25.65
Ours	**23.90**	**27.17**	**29.33**	**27.32**	**19.64**	**19.46**	**29.84**	**24.66**	**25.82**
SSIM(↑)									
C2-Matching [22]	0.777	0.859	0.890	0.888	0.690	0.632	0.944	0.863	0.839
MRefSR [61]	0.783	0.861	0.894	0.895	0.699	0.640	0.947	0.868	0.844
NeRFLiX [66]	0.785	0.862	0.890	0.897	0.698	0.639	0.948	0.870	0.844
Ours	**0.797**	**0.863**	**0.897**	**0.901**	**0.702**	**0.648**	**0.950**	**0.877**	**0.850**
LPIPS(↓)									
C2-Matching [22]	0.201	0.132	0.125	0.133	0.222	0.257	0.087	0.155	0.154
MRefSR [61]	0.191	0.124	0.114	0.124	0.206	**0.245**	0.081	0.147	0.144
NeRFLiX [66]	0.185	**0.123**	0.122	**0.120**	**0.205**	0.250	0.080	0.144	0.143
Ours	**0.184**	0.129	**0.113**	**0.120**	0.208	0.248	**0.079**	**0.141**	**0.142**