Acknowledgments. We thank the anonymous reviewers for their valuable feedback. This work is supported by the Early Career Scheme of the Research Grants Council (grant # 27207224), the HKU-100 Award, and in part by NSF China (No. 62172363). Siyan Dong would also like to thank the support from HKU Musketeers Foundation Institute of Data Science for the Postdoctoral Research Fellowship.

6 Detailed Introduction of Our Dataset

We investigate the problem of performing NeRF-based 3D reconstruction from images with significant pose errors. To benchmark the aforementioned problem, we collect 3D meshes from BlendedMVS [58] and generate a new inward-facing dataset. We select 8 representative scenes and uniformly sample camera viewpoints by different strategies in the hemisphere around each mesh. The statistics and visualization of each scene are shown in Table 4 and Figure 8.

Table 4: Scene statistics of our proposed dataset. For each scene, we report the number of total images, the rotation & translation errors of the camera pose estimates from our SfM module, and the number of accurate poses under different tolerance thresholds.

		Camera pose error (cm, deg)		# < thresh	
Scene	# Images	Mean	Median	1, 1	20, 20
Baby	30	3.94, 59.93	0.28, 0.14	20	20
Bear	45	3.64, 55.24	1.11, 0.56	17	31
Bell	18	10.15, 59.91	0.39, 0.31	12	12
Clock	108	14.17, 52.14	0.80, 0.24	73	76
Deaf	30	9.58, 35.95	1.82, 0.99	1	24
Farmer	18	0.75, 19.88	0.43, 0.33	15	16
Pavilion	18	2.46, 12.07	0.35, 0.26	16	16
Sculpture	30	8.29, 30.76	6.35, 2.62	0	25

Most of the selected scenes contain 18-45 training images, with the exception of the scene *Clock*, which comprises 108 training images. We calculate the initial camera poses for the training images with our SfM module (COLMAP [40] with SuperPoint [12] and SuperGlue [38]). The reconstruction result of each scene contains significant incorrect poses, evident from the gap between the mean and median pose errors. We also report the number of correct poses in each scene, with two sets of pose error thresholds: 1 cm 1 deg, and 20 cm 20 deg.

The incorrect camera pose estimates are primarily due to incorrect keypoint matches. To address this, we have enhanced COLMAP with SuperPoint and SuperGlue to create our SfM module. Through our experiments, we have consistently observed our SfM module outperforming the standard COLMAP. All experiments mentioned in the main paper, conducted on our dataset, utilized the same initial poses generated by our SfM module. An example of a typical

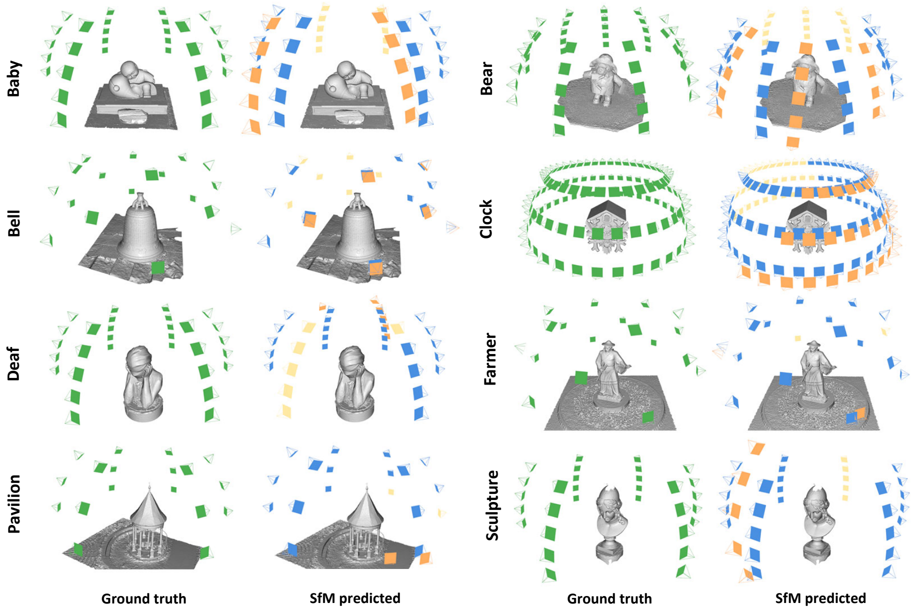

Fig. 8: Illustration of the training viewpoints in our dataset. For each scene, the left image visualizes the ground truth camera poses (green frustums). The right image displays the SfM predicted results, which are utilized as input poses for NeRF training. The blue frustums represent poses within 20 cm and 20 deg. The orange frustums represent those larger than 20 cm and 20 deg, while the yellow frustums denote the ground truth poses of the orange ones.

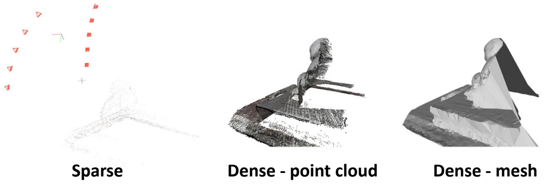

Fig. 9: Illustration of COLMAP [40] results on scene *Baby*. Both sparse and dense reconstruction processes are unsuccessful. Only 1/3 of the viewpoints are estimated with reasonably accurate poses, and about half of the mesh is reconstructed.

COLMAP result is shown in Figure 9. We also showcase the dense reconstruction results with COLMAP MVS [41].

Evaluation metrics. Following the evaluation protocol of previous research [24, 49], we choose Chamfer distance and F-score as metrics to evaluate 3D reconstruction quality on our dataset. As the optimization of camera poses leads to changes in the coordinate system of the scene, it's crucial to align the obtained mesh with ground truth before evaluation. This is achieved by aligning the estimated camera poses with the ground truth. Since our dataset includes outliers that could disturb the alignment process, we manually filter out camera poses with an initial error exceeding 20 cm and 20 degrees during the process. We first align the global orientation on top of the estimated and ground truth camera rotations. Then, we solve a convex optimization problem to find the global scaling factor and translation vector. As a result, we obtain the 7-degree-of-freedom relative transformation matrix between the estimated and ground truth camera poses. After the alignment, we scale the reconstructed and ground truth meshes by a factor of 10, sample $K = 100,000$ points on each mesh surface, and calculate the metrics on top of the sampled points.

To calculate the Chamfer distance, we first compute two measurements: accuracy $Acc(\cdot)$ and completeness $Com(\cdot)$. They are computed by:

$$\begin{aligned} Acc(S_{rec}, S_{gt}) &= \frac{1}{K} \sum_{p \in S_{rec}} \min_{q \in S_{gt}} ||p - q||_1, \\ Com(S_{rec}, S_{gt}) &= \frac{1}{K} \sum_{q \in S_{gt}} \min_{p \in S_{rec}} ||p - q||_1. \end{aligned} \tag{7}$$

Then, the Chamfer distance is calculated as the mean of the two aforementioned measurements:

$$CD(S_{rec}, S_{gt}) = \frac{Acc(S_{rec}, S_{gt}) + Com(S_{rec}, S_{gt})}{2}. \tag{8}$$

To calculate the F-score, we compute two measurements: precision $Pre(\cdot)$ and recall $Rec(\cdot)$. They are computed by:

$$\begin{aligned} Pre(S_{rec}, S_{gt}) &= \frac{1}{K} \sum_{p \in S_{rec}} \max_{q \in S_{gt}} ||p - q||_0, \\ Rec(S_{rec}, S_{gt}) &= \frac{1}{K} \sum_{q \in S_{gt}} \max_{p \in S_{rec}} ||p - q||_0, \end{aligned} \tag{9}$$

where $|| \cdot ||_0$ indicates inlier/outlier points with a distance threshold d:

$$|| \cdot ||_0 = \begin{cases} 1 & || \cdot ||_1 < d \\ 0 & otherwise \end{cases} \tag{10}$$

We set $d = 0.64$. The F-score is calculated as the harmonic mean of the precision and recall:

$$F - score(S_{rec}, S_{gt}) = \frac{2 \cdot Pre(S_{rec}, S_{gt}) \cdot Rec(S_{rec}, S_{gt})}{Pre(S_{rec}, S_{gt}) + Rec(S_{rec}, S_{gt})}. \tag{11}$$

7 More Evaluations

Evaluation of camera pose errors. In Table 5, we report mean pose errors on our dataset. To ensure a fair comparison, we use the confidence-weighted mean error, denoted as SG-W. This is because not all poses have an equal impact on our method. SG-W achieves substantial error reduction compared to competitors. We also report the results with hard outlier rejection by considering only the selected training viewpoints in the final epoch, denoted by SG-H. The decrease in pose errors is more significant, and it achieves mean precision of 68% and recall of 80% for outlier rejection. Here, the ground truth outliers are defined by errors $>$ 20 cm 20 deg of their initial camera poses.

Table 5: Evaluation of camera pose errors on the proposed dataset.

Mean errors	BARF	SCNeRF	GARF	L2G-NeRF	Joint-TensoRF	CamP	PoRF	SG	SG-W	SG-H
Translation (cm)	0.71	0.69	0.78	0.82	0.88	0.69	0.75	0.70	0.45	0.09
Rotation (deg)	39.40	40.70	47.07	45.31	46.55	40.82	43.26	39.95	25.73	4.62

Comparison with recent camera pose optimizer CamP [35]. We have tested CamP in the scene *Bell*. The mean camera pose error is 1.01 cm 60.04 deg. We find that CamP does not always guarantee the prevention of the optimization falling into local minima, especially when there are outlier camera poses. As a result, the pose errors are similar to those of SCNeRF.

Additional qualitative comparison on our dataset. Due to the page number limit, we show qualitative results for only 4 scenes in Figure 1 and Figure 5 in the main paper. The remaining results can be found in Figure 10.

Discussion and comparison with SPARF [47]. **Problem setting**: SPARF tackles the problem of few-shot reconstruction (typically 3 views) with moderate camera poses; we focus on the common inward-facing scenario that can include real large pose errors. **Method design**: While both SPARF and our method utilize multi-view keypoint correspondences, we introduce an IoU formulation. This allows us to enhance the implicit geometry in a more continuous optimization space (i.e., aligning MoGs), as opposed to directly applying the conventional re-projection error. Additionally, we do not require dense depth rendering, which

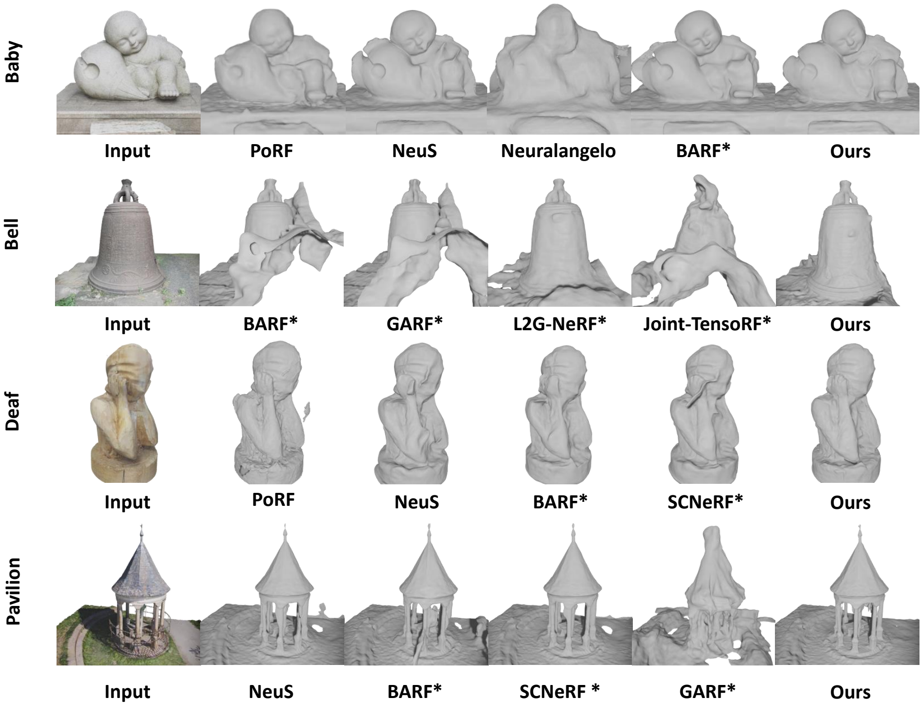

Fig. 10: Additional qualitative comparison on our dataset.

can be computationally expensive and consume a significant amount of memory. **Quantitative comparison**: We have tested SPARF on the scene *Bell*. Through the experiment, we find that the outliers' matches and poses can significantly mislead SPARF's optimization process. This results in a mean pose error of 1.17 cm 165.19 deg (SG-W achieves 0.85 cm and 43.09 deg).

Upper bound. To better understand the reconstruction results on our dataset, we provide an upper bound estimation on top of our method. We utilize the ground truth camera poses and set these poses as fixed. The results are shown in Table 6. Our method has demonstrated significant improvements compared to plain NeuS [49]. However, there is still room for further research to enhance performance. Although our approach, which involves the soft rejection of outliers, mitigates the influence of these outliers, it doesn't completely eliminate them. These outliers can still affect the reconstruction performance. We believe that combining our method with hard outlier rejection or visual (re)localization techniques could potentially address this issue. This presents a promising direction for future research.

Neuralangelo with pose optimization. In the main paper, we utilize NeuS [49] as the backbone for 3D reconstruction and compare various pose optimization

Table 6: Additional quantitative results on our dataset.

		Baby	Bear	Bell	Clock	Deaf	Farmer	Pavilion	Sculpture	Mean
Chamfer distance ↓	CasMVSNet [18]	1.04	0.78	1.93	1.24	1.04	3.25	1.16	1.12	1.45
	IterMVS [48]	0.73	0.65	1.51	0.90	0.86	1.70	0.90	0.57	0.98
	BARF [25]#	0.72	0.40	5.11	0.26	0.49	5.73	5.29	0.68	2.34
	SCNeRF [22]#	0.71	0.64	-	0.49	0.50	5.18	2.60	0.58	1.53†
	GARF [8]#	-	0.47	-	2.21	0.89	5.37	-	0.59	1.91†
	L2G-NeRF [6]#	0.74	1.85	6.82	0.35	0.53	6.04	3.12	2.45	2.74
	Joint-TensoRF [7]#	-	1.34	-	0.73	1.30	6.23	3.35	1.45	2.40†
	SG-NeRF (Ours)	0.56	0.25	0.98	0.15	0.45	0.87	0.20	0.22	0.46
	Upper bound	0.37	0.08	0.23	0.08	0.31	0.19	0.14	0.16	0.20
F-score ↑	CasMVSNet [18]	0.43	0.57	0.33	0.39	0.40	0.27	0.51	0.46	0.42
	IterMVS [48]	0.53	0.65	0.39	0.53	0.41	0.38	0.57	0.62	0.51
	BARF [25]#	0.67	0.87	0.08	0.90	0.76	0.10	0.41	0.81	0.58
	SCNeRF [22]#	0.54	0.79	-	0.78	0.77	0.12	0.37	0.69	0.58†
	GARF [8]#	-	0.83	-	0.11	0.46	0.12	-	0.80	0.46†
	L2G-NeRF [6]#	0.68	0.65	0.02	0.86	0.79	0.10	0.22	0.16	0.44
	Joint-TensoRF [7]#	-	0.35	-	0.57	0.22	0.08	0.15	0.25	0.27†
	SG-NeRF (Ours)	0.74	0.93	0.71	0.96	0.87	0.76	0.94	0.92	0.85
	Upper bound	0.80	0.99	0.92	0.99	0.91	0.95	0.97	0.94	0.93

methods. In table 6, we present the results of combining the pose optimization methods with Neuralangelo [24]. The methods are identified as BARF [25]#, SCNeRF [22]#, and so on. As observed, the results are worse than those with NeuS, since Neuralangelo is more sensitive to outliers. The results further confirm our effectiveness in handling outliers. They also demonstrate the non-trivial design for joint optimization of the radiance field and camera poses, instead of simply combining different methods.

Comparison with MVS methods. Since our focus is on neural surface reconstruction with radiance fields, we do not compare our work with classical Multi-View Stereo (MVS) algorithms in the main paper. Here, in table 6, we report the results from two recent MVS methods: CasMVSNet [18] and IterMVS [48]. As MVS-based methods generally utilize the epipolar geometry prior that heavily relies on camera poses, these methods are sensitive to pose errors.

Additional comparison with more surface reconstruction methods. Since our goal is to mitigate the impact of outliers, we primarily compare with pose optimization methods in the main paper. Here, we also run HF-NeuS [51], Voxsurf [54], and NeuDA [5] on the scene *Bell*. The Chamfer distances (1.41, 2.64, and 1.72) are larger than our method. We notice that these methods that aim to reconstruct details tend to be sensitive to large pose errors.

Bundle adjustment (BA) with the pruned scene graph. We have tried to run BA on the pruned graph, but the pose updates are minimal. Note that the removed edges by our pruning step were considered good and consistent during previous COLMAP processing (including BA). Since BA does not use residuals on dense pixel colors, the new round of BA on the sparse graph makes only a small contribution. As a result, NeuS still produces unsatisfactory results.

The choice of λ. The parameter λ adjusts the impact of PSNR on the CS update. In Table 7, we report the quantitative results on our dataset with various λ values. The results show that as λ increases, the performance first improves and then decreases. This suggests selecting an appropriate λ that balances the sparse (keypoint matches) and dense (photometric residuals) information. Thus, we set $\lambda = 1$.

Table 7: Results with different λ.

λ	0.2	0.5	1.0	1.5	2.0	3.0
Camfer ↓	0.5058	0.4820	**0.4646**	0.4788	0.4885	0.4818
F-score ↑	0.8379	0.8516	**0.8524**	0.8461	0.8438	0.8384

8 Real-World Scene Reconstruction

In this paper, we tackle a challenging, yet practical scenario where images are casually captured without careful selection. To highlight the benefits of our method under such conditions, we showcase a real-world reconstruction example. We casually capture 20 images of a toy object in an inward-facing manner. While COLMAP fails to solve the camera poses for this scene, our SfM module estimates the poses reasonably. Given the initial camera poses, we train NeuS [49] and our method and then extract meshes for a quantitative comparison. Figure 11 illustrates the results. We can observe that NeuS's mesh contains wrong geometries and tends to be over-smoothed. On the contrary, our reconstruction presents correct structures and more details, demonstrating our robustness to wrong poses.

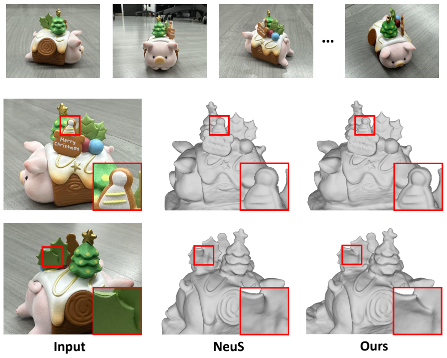

Fig. 11: Qualitative comparison of a real-world scene reconstruction.

References

1. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5470–5479 (2022)

2. Bian, J.W., Bian, W., Prisacariu, V.A., Torr, P.: Porf: Pose residual field for accurate neural surface reconstruction. In: ICLR (2024)
3. Bian, W., Wang, Z., Li, K., Bian, J.W., Prisacariu, V.A.: Nope-nerf: Optimising neural radiance field with no pose prior. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4160–4169 (2023)
4. Brachmann, E., Rother, C.: Visual camera re-localization from rgb and rgb-d images using dsac. IEEE transactions on pattern analysis and machine intelligence **44**(9), 5847–5865 (2021)
5. Cai, B., Huang, J., Jia, R., Lv, C., Fu, H.: Neuda: Neural deformable anchor for high-fidelity implicit surface reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8476–8485 (2023)
6. Chen, Y., Chen, X., Wang, X., Zhang, Q., Guo, Y., Shan, Y., Wang, F.: Local-to-global registration for bundle-adjusting neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8264–8273 (2023)
7. Cheng, B.Y., Chiu, W.C., Liu, Y.L.: Improving robustness for joint optimization of camera poses and decomposed low-rank tensorial radiance fields. arXiv preprint arXiv:2402.13252 (2024)
8. Chng, S.F., Ramasinghe, S., Sherrah, J., Lucey, S.: Gaussian activated neural radiance fields for high fidelity reconstruction and pose estimation. In: European Conference on Computer Vision. pp. 264–280. Springer (2022)
9. Cui, Z., Tan, P.: Global structure-from-motion by similarity averaging. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 864–872 (2015)
10. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. pp. 303–312 (1996)
11. 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 (CVPR) (June 2022)
12. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 224–236 (2018)
13. Dong, S., Fan, Q., Wang, H., Shi, J., Yi, L., Funkhouser, T., Chen, B., Guibas, L.J.: Robust neural routing through space partitions for camera relocalization in dynamic indoor environments. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8544–8554 (2021)
14. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24**(6), 381–395 (1981)
15. Frahm, J.M., Pollefeys, M.: Ransac for (quasi-) degenerate data (qdegsac). In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). vol. 1, pp. 453–460. IEEE (2006)
16. Fu, Q., Xu, Q., Ong, Y.S., Tao, W.: Geo-neus: Geometry-consistent neural implicit surfaces learning for multi-view reconstruction. Advances in Neural Information Processing Systems **35**, 3403–3416 (2022)
17. Furukawa, Y., Hernández, C., et al.: Multi-view stereo: A tutorial. Foundations and Trends® in Computer Graphics and Vision **9**(1-2), 1–148 (2015)
18. Gu, X., Fan, Z., Zhu, S., Dai, Z., Tan, F., Tan, P.: Cascade cost volume for high-resolution multi-view stereo and stereo matching. In: Proceedings of the

IEEE/CVF conference on computer vision and pattern recognition. pp. 2495–2504 (2020)

19. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge university press (2003)
20. Heo, H., Kim, T., Lee, J., Lee, J., Kim, S., Kim, H.J., Kim, J.H.: Robust camera pose refinement for multi-resolution hash encoding. arXiv preprint arXiv:2302.01571 (2023)
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. Jeong, Y., Ahn, S., Choy, C., Anandkumar, A., Cho, M., Park, J.: Self-calibrating neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5846–5854 (2021)
23. Lepetit, V., Moreno-Noguer, F., Fua, P.: Ep n p: An accurate o (n) solution to the p n p problem. International journal of computer vision **81**, 155–166 (2009)
24. Li, Z., Müller, T., Evans, A., Taylor, R.H., Unberath, M., Liu, M.Y., Lin, C.H.: Neuralangelo: High-fidelity neural surface reconstruction. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
25. Lin, C.H., Ma, W.C., Torralba, A., Lucey, S.: Barf: Bundle-adjusting neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5741–5751 (2021)
26. Lindenberger, P., Sarlin, P.E., Larsson, V., Pollefeys, M.: Pixel-Perfect Structure-from-Motion with Featuremetric Refinement. In: ICCV (2021)
27. Liu, S., Yu, Y., Pautrat, R., Pollefeys, M., Larsson, V.: 3d line mapping revisited. In: Computer Vision and Pattern Recognition (CVPR) (2023)
28. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: Seminal graphics: pioneering efforts that shaped the field, pp. 347–353 (1998)
29. 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: ECCV (2020)
30. Moreau, A., Piasco, N., Bennehar, M., Tsishkou, D., Stanciulescu, B., de La Fortelle, A.: Crossfire: Camera relocalization on self-supervised features from an implicit representation. arXiv preprint arXiv:2303.04869 (2023)
31. Moreau, A., Piasco, N., Tsishkou, D., Stanciulescu, B., de La Fortelle, A.: Lens: Localization enhanced by nerf synthesis. In: Conference on Robot Learning. pp. 1347–1356. PMLR (2022)
32. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Graph. **41**(4), 102:1–102:15 (Jul 2022). `https://doi.org/10.1145/3528223.3530127`, `https://doi.org/10.1145/3528223.3530127`
33. Oechsle, M., Peng, S., Geiger, A.: Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In: International Conference on Computer Vision (ICCV) (2021)
34. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 165–174 (2019)
35. Park, K., Henzler, P., Mildenhall, B., Barron, J.T., Martin-Brualla, R.: Camp: Camera preconditioning for neural radiance fields. ACM Transactions on Graphics (TOG) **42**(6), 1–11 (2023)

36. 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)
37. Sarlin, P.E., Cadena, C., Siegwart, R., Dymczyk, M.: From coarse to fine: Robust hierarchical localization at large scale. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12716–12725 (2019)
38. Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperGlue: Learning feature matching with graph neural networks. In: CVPR (2020)
39. Sattler, T., Leibe, B., Kobbelt, L.: Efficient & effective prioritized matching for large-scale image-based localization. IEEE transactions on pattern analysis and machine intelligence **39**(9), 1744–1756 (2016)
40. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
41. Schönberger, J.L., Zheng, E., Pollefeys, M., Frahm, J.M.: Pixelwise view selection for unstructured multi-view stereo. In: European Conference on Computer Vision (ECCV) (2016)
42. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3d. In: ACM siggraph 2006 papers, pp. 835–846 (2006)
43. Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the world from internet photo collections. International journal of computer vision **80**, 189–210 (2008)
44. Sweeney, C.: Theia multiview geometry library: Tutorial & reference. `http://theia-sfm.org`
45. Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., Kristoffersen, A., Austin, J., Salahi, K., Ahuja, A., et al.: Nerfstudio: A modular framework for neural radiance field development. In: ACM SIGGRAPH 2023 Conference Proceedings. pp. 1–12 (2023)
46. Torr, P.H.: An assessment of information criteria for motion model selection. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 47–52. IEEE (1997)
47. Truong, P., Rakotosaona, M.J., Manhardt, F., Tombari, F.: Sparf: Neural radiance fields from sparse and noisy poses. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4190–4200 (2023)
48. Wang, F., Galliani, S., Vogel, C., Pollefeys, M.: Itermvs: Iterative probability estimation for efficient multi-view stereo. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8606–8615 (2022)
49. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. NeurIPS (2021)
50. Wang, Y., Han, Q., Habermann, M., Daniilidis, K., Theobalt, C., Liu, L.: Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3295–3306 (2023)
51. Wang, Y., Skorokhodov, I., Wonka, P.: Hf-neus: Improved surface reconstruction using high-frequency details. Advances in Neural Information Processing Systems **35**, 1966–1978 (2022)
52. Wang, Z., Wu, S., Xie, W., Chen, M., Prisacariu, V.A.: Nerf–: Neural radiance fields without known camera parameters. arXiv preprint arXiv:2102.07064 (2021)
53. Wu, C.: Visualsfm: A visual structure from motion system. http://www. cs. washington. edu/homes/ccwu/vsfm (2011)

54. Wu, T., Wang, J., Pan, X., Xu, X., Theobalt, C., Liu, Z., Lin, D.: Voxurf: Voxel-based efficient and accurate neural surface reconstruction. arXiv preprint arXiv:2208.12697 (2022)
55. Xu, X., Yang, Y., Mo, K., Pan, B., Yi, L., Guibas, L.: Jacobinerf: Nerf shaping with mutual information gradients. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16498–16507 (2023)
56. Yan, Q., Wang, Q., Zhao, K., Chen, J., Li, B., Chu, X., Deng, F.: Cf-nerf: Camera parameter free neural radiance fields with incremental learning. arXiv preprint arXiv:2312.08760 (2023)
57. Yao, Y., Luo, Z., Li, S., Fang, T., Quan, L.: Mvsnet: Depth inference for unstructured multi-view stereo. European Conference on Computer Vision (ECCV) (2018)
58. Yao, Y., Luo, Z., Li, S., Zhang, J., Ren, Y., Zhou, L., Fang, T., Quan, L.: Blendedmvs: A large-scale dataset for generalized multi-view stereo networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1790–1799 (2020)
59. Yariv, L., Gu, J., Kasten, Y., Lipman, Y.: Volume rendering of neural implicit surfaces. In: Thirty-Fifth Conference on Neural Information Processing Systems (2021)
60. Yariv, L., Kasten, Y., Moran, D., Galun, M., Atzmon, M., Ronen, B., Lipman, Y.: Multiview neural surface reconstruction by disentangling geometry and appearance. Advances in Neural Information Processing Systems **33** (2020)
61. Zhang, K., Riegler, G., Snavely, N., Koltun, V.: Nerf++: Analyzing and improving neural radiance fields. arXiv:2010.07492 (2020)
62. Zhu, B., Yang, Y., Wang, X., Zheng, Y., Guibas, L.: Vdn-nerf: Resolving shape-radiance ambiguity via view-dependence normalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 35–45 (2023)
63. Zhu, Z., Peng, S., Larsson, V., Xu, W., Bao, H., Cui, Z., Oswald, M.R., Pollefeys, M.: Nice-slam: Neural implicit scalable encoding for slam. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12786–12796 (2022)