Supplementary material for
Nerfbusters: Removing Ghostly Artifacts from Casually Captured NeRFs

Frederik Warburg* Ethan Weber* Matthew Tancik
Aleksander Holynski Angjoo Kanazawa

University of California, Berkeley

	Translation to NN	Rotation to NN
Nerfbusters	0.62	28.51
Synthetic [4]	0.29	18.12
LLFF [3]	0.24	2.07
Phototourism [2]	0.01	9.47
MipNeRF-360 [1]	0.07	10.49

Table 1: Average translation and rotation difference between eval poses and closest training poses (nearest neighbor (NN)). The proposed Nerfbusters dataset is larger in translation and rotation differences than the current popular NeRF datasets.

1. Supplement

In this document, we seek to provide more details about the proposed dataset, evaluation protocol, and per-scene results.

2. Dataset

We compare the proposed dataset with popular NeRF datasets. We first normalize each scene's camera poses (train and evaluation cameras combined) to the range [-1, 1]. We then find the nearest training image for each evaluation image. We compute the translation and rotation difference between the nearest neighbors (NN) and average over the number of evaluation poses. Tab. 1 shows that the proposed Nerfbusters dataset has a larger translation and rotation difference between training and evaluation frames than existing popular NeRF datasets.

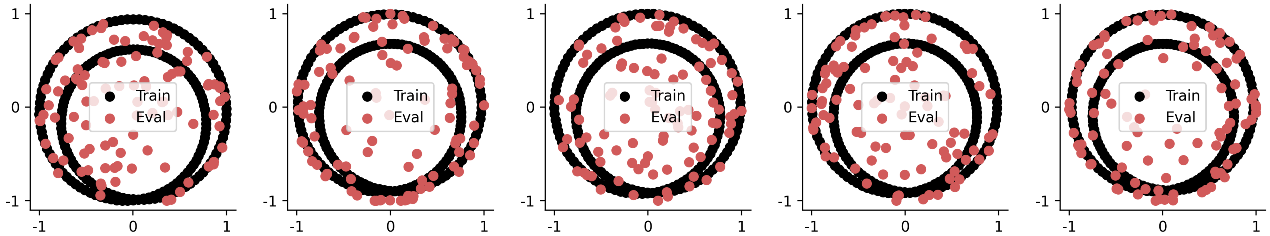
Tab. 1 compares translation and rotation differences between training and evaluation images. Although this is a useful approximation of the "difficulty" of the datasets, it does not encompass variations in the camera intrinsics, e.g. we can obtain a similar effect as a translation by changing the focal length. In practice though, a common assumption is that the same camera intrinsics are used for training and test renderings. This assumption does not hold for the Phototourism [2] dataset, where the camera intrinsics vary for each training frame. Therefore, Tab. 1 is a crude approximation of difficulty for this dataset.

Figs. 1 to 5 show top-down views of training and evaluation views for each scene in the Synthetic [4], LLFF [3], MipNeRF-360 [1], Phototourism [2], and our Nerfbusters dataset. Most of the datasets use every 8th frame for evaluation, resulting in training and evaluation views being very close. The train and test frames are further apart in the Nerfbusters dataset, faithfully representing some of the real-world challenges that users of NeRF experience when rendering fly-throughs of their captured scenes.

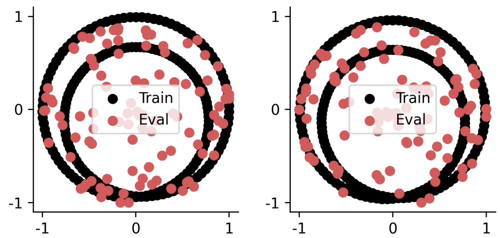
3. Per-Scene Evaluation

We provide per-scene evaluation for our method and the baseline methods in Tabs. 2 to 13. The average over the 12 datasets is included in the main paper.

*Denotes equal contribution

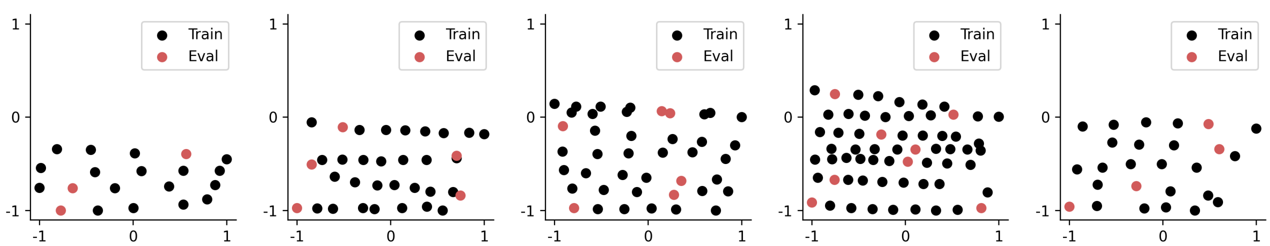


(a) Drum (b) Ficus (c) Hotdog (d) Lego (e) Materials

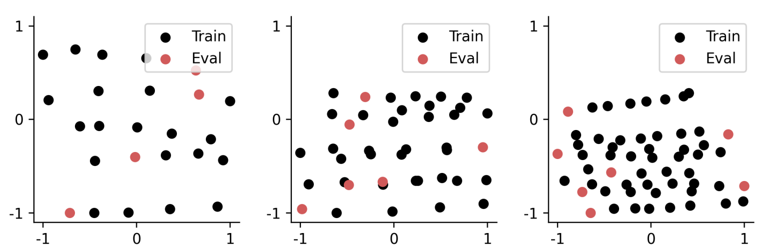


(f) Mic (g) Ship

Figure 1: **Synethic NeRF.**

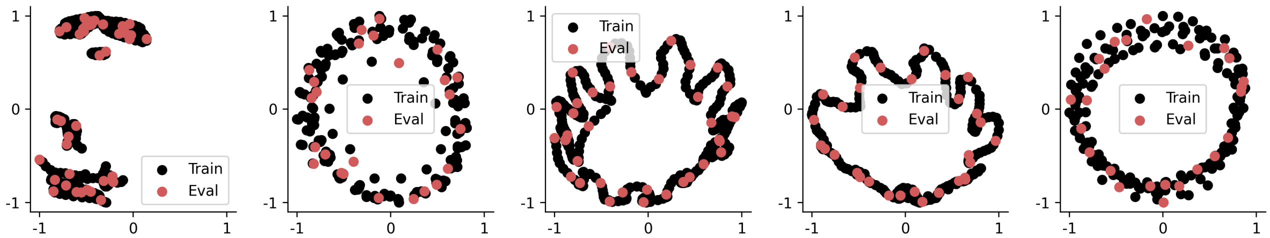


(a) Fern (b) Flower (c) Fortress (d) Horns (e) Leaves

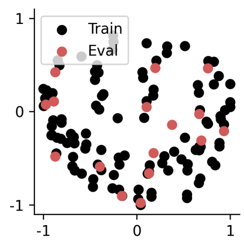


(f) Orchids (g) Room (h) Trex

Figure 2: **LLFF dataset.**

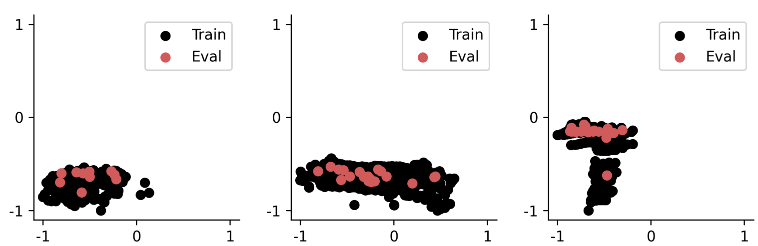


(a) Room (b) Bicycle (c) Bonsai (d) Counter (e) Garden



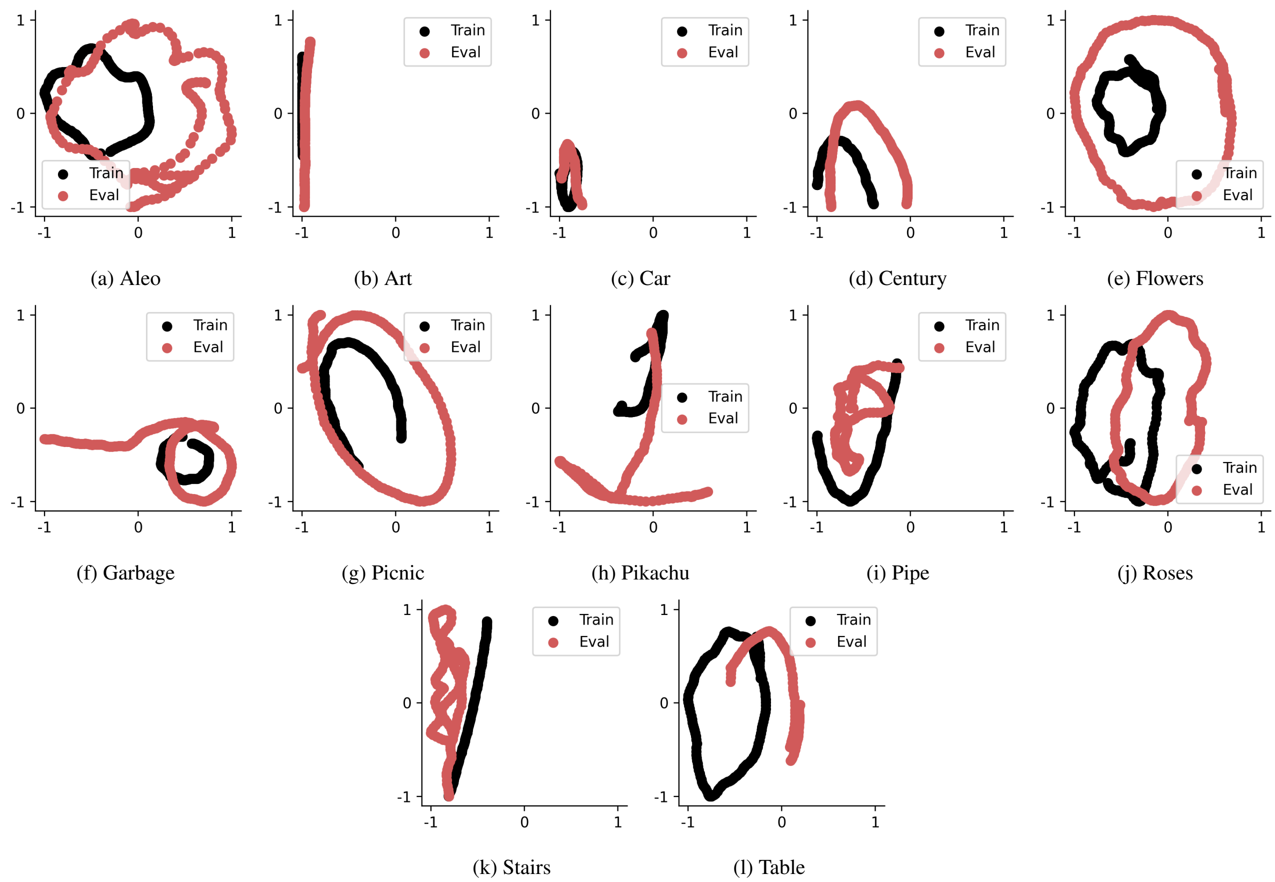
(f) Stump

Figure 3: **MipNeRF 360 dataset.**



(a) Brandenburg (b) Trevi (c) Sacre

Figure 4: **Phototourism dataset.**



(a) Aleo (b) Art (c) Car (d) Century (e) Flowers

(f) Garbage (g) Picnic (h) Pikachu (i) Pipe (j) Roses

(k) Stairs (l) Table

Figure 5: **Nerfbusters dataset.**

	PSNR ↑	SSIM ↑	LPIPS ↓	Depth ↓	Disp. ↓	Mean ° ↓	Median ° ↓	% 30° ↑	Coverage ↑
Nerfacto Pseudo GT	24.02	0.8591	0.0801	0.00	0.00	0.01	0.00	1.00	0.84
Nerfacto	18.33	0.6037	0.2968	0.01	0.04	45.50	39.01	0.37	0.84
+ Visibility Loss	18.32	0.6063	0.2929	0.01	0.02	45.28	38.81	0.37	0.84
+ Vis + Sparsity	18.41	0.6082	0.2895	0.01	0.02	45.19	38.76	0.37	0.84
+ Vis + TV	18.31	0.6073	0.2983	0.01	0.02	46.12	39.34	0.36	0.84
+ Vis + RegNeRF	18.31	0.6081	0.2933	0.01	0.02	45.44	39.04	0.37	0.84
+ Vis + DSDS (Ours)	18.75	0.6390	0.2441	0.01	0.02	45.43	39.14	0.36	0.80

Table 2: **"aloe" capture quantitative evaluation.**

	PSNR ↑	SSIM ↑	LPIPS ↓	Depth ↓	Disp. ↓	Mean ° ↓	Median ° ↓	% 30° ↑	Coverage ↑
Nerfacto Pseudo GT	25.25	0.8534	0.0710	0.00	0.00	0.01	0.00	1.00	0.88
Nerfacto	14.72	0.3784	0.3978	1025.94	0.20	65.26	60.02	0.20	0.87
+ Visibility Loss	14.77	0.3860	0.3933	1006.97	0.17	64.55	59.30	0.20	0.86
+ Vis + Sparsity	14.86	0.3900	0.3901	907.83	0.17	64.48	59.19	0.20	0.86
+ Vis + TV	15.02	0.3991	0.3838	719.83	0.16	68.80	63.99	0.17	0.85
+ Vis + RegNeRF	13.30	0.2995	0.4647	1386.03	2.57	72.13	67.93	0.15	0.89
+ Vis + DSDS (Ours)	15.05	0.4336	0.3408	390.80	0.14	62.21	56.89	0.21	0.74

Table 3: **"art" capture quantitative evaluation.**

	PSNR ↑	SSIM ↑	LPIPS ↓	Depth ↓	Disp. ↓	Mean ° ↓	Median ° ↓	% 30° ↑	Coverage ↑
Nerfacto Pseudo GT	22.98	0.7480	0.2093	0.00	0.00	0.01	0.00	1.00	0.80
Nerfacto	16.57	0.4630	0.4090	240.77	0.56	70.51	66.31	0.17	0.81
+ Visibility Loss	16.67	0.4609	0.4136	133.80	0.51	70.19	65.94	0.18	0.81
+ Vis + Sparsity	16.70	0.4632	0.4149	137.97	0.38	70.57	66.38	0.17	0.81
+ Vis + TV	16.87	0.4827	0.3956	116.45	0.39	73.52	69.63	0.16	0.78
+ Vis + RegNeRF	16.62	0.4654	0.4099	745.14	0.56	70.73	66.59	0.17	0.82
+ Vis + DSDS (Ours)	17.40	0.5898	0.2219	23.76	0.14	65.99	60.43	0.21	0.55

Table 4: **"car" capture quantitative evaluation.**

	PSNR ↑	SSIM ↑	LPIPS ↓	Depth ↓	Disp. ↓	Mean ° ↓	Median ° ↓	% 30° ↑	Coverage ↑
Nerfacto Pseudo GT	23.76	0.8699	0.0790	0.00	0.00	0.01	0.00	1.00	0.95
Nerfacto	13.97	0.3938	0.4998	2.27	3.34	65.46	59.70	0.22	0.97
+ Visibility Loss	13.90	0.4022	0.4858	1.27	3.50	63.58	57.34	0.24	0.96
+ Vis + Sparsity	13.91	0.4021	0.4899	1.27	2.35	63.44	57.26	0.24	0.96
+ Vis + TV	14.25	0.4286	0.4464	1.03	1.40	69.38	64.59	0.18	0.96
+ Vis + RegNeRF	12.63	0.3459	0.5672	2.84	3.95	71.88	67.44	0.18	0.97
+ Vis + DSDS (Ours)	15.01	0.5028	0.2840	0.41	0.35	59.15	52.75	0.25	0.72

Table 5: **"century" capture quantitative evaluation.**

	PSNR ↑	SSIM ↑	LPIPS ↓	Depth ↓	Disp. ↓	Mean ° ↓	Median ° ↓	% 30° ↑	Coverage ↑
Nerfacto Pseudo GT	26.67	0.8687	0.1044	0.00	0.00	0.01	0.00	1.00	0.97
Nerfacto	15.68	0.4654	0.4413	14.28	0.22	65.45	60.02	0.20	0.93
+ Visibility Loss	15.77	0.4691	0.4322	17.25	0.12	65.35	59.92	0.21	0.93
+ Vis + Sparsity	15.67	0.4667	0.4309	17.21	0.12	65.31	59.87	0.21	0.93
+ Vis + TV	15.47	0.4718	0.4314	13.03	0.11	68.00	62.88	0.19	0.91
+ Vis + RegNeRF	15.67	0.4649	0.4356	13.85	0.12	65.68	60.26	0.20	0.93
+ Vis + DSDS (Ours)	15.52	0.5086	0.3156	5.77	0.08	61.18	54.47	0.24	0.62

Table 6: **"flowers" capture quantitative evaluation.**

	PSNR ↑	SSIM ↑	LPIPS ↓	Depth ↓	Disp. ↓	Mean ° ↓	Median ° ↓	% 30° ↑	Coverage ↑
Nerfacto Pseudo GT	24.09	0.8151	0.1424	0.00	0.00	0.01	0.00	1.00	1.00
Nerfacto	14.86	0.4047	0.5072	0.09	10.81	63.75	58.69	0.25	1.00
+ Visibility Loss	14.91	0.4197	0.4973	0.02	6.65	62.67	57.31	0.25	1.00
+ Vis + Sparsity	15.04	0.4193	0.4977	0.02	9.29	62.27	56.90	0.26	1.00
+ Vis + TV	15.20	0.4212	0.5110	0.04	1.18	73.74	70.17	0.15	1.00
+ Vis + RegNeRF	15.13	0.4205	0.4926	0.03	5.42	63.18	57.80	0.25	1.00
+ Vis + DSDS (Ours)	15.86	0.4466	0.3726	0.00	0.11	53.09	45.57	0.31	0.63

Table 7: **"garbage" capture quantitative evaluation.**

	PSNR ↑	SSIM ↑	LPIPS ↓	Depth ↓	Disp. ↓	Mean ° ↓	Median ° ↓	% 30° ↑	Coverage ↑
Nerfacto Pseudo GT	23.44	0.7616	0.1774	0.00	0.00	0.01	0.00	1.00	0.95
Nerfacto	15.99	0.3123	0.5019	0.12	1.17	61.20	55.39	0.24	0.96
+ Visibility Loss	15.93	0.3117	0.4992	0.11	0.90	60.57	54.66	0.25	0.96
+ Vis + Sparsity	15.97	0.3145	0.4992	0.10	0.77	60.91	55.02	0.25	0.96
+ Vis + TV	16.03	0.3168	0.5123	0.11	0.76	68.94	63.99	0.19	0.96
+ Vis + RegNeRF	15.77	0.3014	0.5149	0.14	1.12	62.16	56.47	0.23	0.96
+ Vis + DSDS (Ours)	15.72	0.4560	0.3074	0.11	0.12	54.32	46.71	0.31	0.58

Table 8: **"picnic" capture quantitative evaluation.**

	PSNR ↑	SSIM ↑	LPIPS ↓	Depth ↓	Disp. ↓	Mean ° ↓	Median ° ↓	% 30° ↑	Coverage ↑
Nerfacto Pseudo GT	31.69	0.9606	0.0399	0.00	0.00	0.01	0.00	1.00	0.92
Nerfacto	20.31	0.6903	0.3267	0.60	0.83	56.86	48.00	0.29	0.93
+ Visibility Loss	25.83	0.8837	0.0874	0.01	0.01	42.55	34.20	0.44	0.77
+ Vis + Sparsity	25.79	0.8806	0.0909	0.00	0.01	42.59	34.28	0.44	0.77
+ Vis + TV	26.23	0.9000	0.0867	0.00	0.01	47.29	40.29	0.36	0.71
+ Vis + RegNeRF	25.24	0.8734	0.0970	0.01	0.01	42.62	34.26	0.44	0.77
+ Vis + DSDS (Ours)	25.71	0.9048	0.0502	0.00	0.00	45.90	39.17	0.36	0.28

Table 9: **"pikachu" capture quantitative evaluation.**

	PSNR ↑	SSIM ↑	LPIPS ↓	Depth ↓	Disp. ↓	Mean ° ↓	Median ° ↓	% 30° ↑	Coverage ↑
Nerfacto Pseudo GT	23.38	0.8008	0.1384	0.00	0.00	0.01	0.00	1.00	0.93
Nerfacto	19.62	0.5910	0.2673	0.08	0.07	59.62	53.54	0.25	0.93
+ Visibility Loss	19.61	0.5914	0.2637	0.08	0.07	59.27	53.12	0.25	0.93
+ Vis + Sparsity	19.67	0.5906	0.2697	0.07	0.07	60.28	54.29	0.24	0.93
+ Vis + TV	19.64	0.5891	0.2793	0.07	0.07	65.09	59.67	0.20	0.93
+ Vis + RegNeRF	19.62	0.5917	0.2665	0.15	0.07	59.46	53.35	0.25	0.93
+ Vis + DSDS (Ours)	19.23	0.6165	0.2387	0.08	0.08	58.58	52.35	0.26	0.82

Table 10: **"pipe" capture quantitative evaluation.**

	PSNR ↑	SSIM ↑	LPIPS ↓	Depth ↓	Disp. ↓	Mean ° ↓	Median ° ↓	% 30° ↑	Coverage ↑
Nerfacto Pseudo GT	29.66	0.9349	0.0476	0.00	0.00	0.01	0.00	1.00	0.57
Nerfacto	16.42	0.6146	0.3479	229.88	0.38	74.54	70.76	0.17	0.71
+ Visibility Loss	20.55	0.7090	0.1824	40.48	0.05	60.30	52.21	0.28	0.39
+ Vis + Sparsity	20.48	0.7066	0.1867	40.85	0.05	59.98	51.79	0.28	0.39
+ Vis + TV	20.61	0.7148	0.1723	36.89	0.05	60.68	52.63	0.27	0.39
+ Vis + RegNeRF	20.37	0.7038	0.1866	40.52	0.05	60.35	52.25	0.28	0.39
+ Vis + DSDS (Ours)	20.20	0.7535	0.1254	232.09	0.03	55.40	46.52	0.32	0.25

Table 11: **"plant" capture quantitative evaluation.**

	PSNR ↑	SSIM ↑	LPIPS ↓	Depth ↓	Disp. ↓	Mean ° ↓	Median ° ↓	% 30° ↑	Coverage ↑
Nerfacto Pseudo GT	29.51	0.9186	0.0554	0.00	0.00	0.01	0.00	1.00	0.91
Nerfacto	19.96	0.7159	0.2383	1.23	0.02	47.34	40.24	0.36	0.92
+ Visibility Loss	19.94	0.7173	0.2412	0.64	0.02	46.67	39.59	0.37	0.92
+ Vis + Sparsity	19.92	0.7195	0.2416	0.64	0.02	46.34	39.34	0.37	0.92
+ Vis + TV	19.94	0.7190	0.2428	0.69	0.02	47.26	40.16	0.36	0.92
+ Vis + RegNeRF	19.83	0.7168	0.2424	0.61	0.02	47.21	40.14	0.36	0.92
+ Vis + DSDS (Ours)	19.14	0.7149	0.2076	0.37	0.03	47.27	40.45	0.35	0.87

Table 12: **"roses" capture quantitative evaluation.**

	PSNR ↑	SSIM ↑	LPIPS ↓	Depth ↓	Disp. ↓	Mean ° ↓	Median ° ↓	% 30° ↑	Coverage ↑
Nerfacto Pseudo GT	27.30	0.9180	0.0784	0.00	0.00	0.01	0.00	1.00	0.99
Nerfacto	17.60	0.6877	0.3259	0.04	0.47	52.08	43.98	0.33	0.89
+ Visibility Loss	17.51	0.6878	0.3294	0.04	0.48	51.82	43.62	0.34	0.88
+ Vis + Sparsity	17.35	0.6818	0.3330	0.04	0.48	51.86	43.71	0.33	0.88
+ Vis + TV	16.49	0.6895	0.3311	0.04	0.41	54.34	46.63	0.31	0.86
+ Vis + RegNeRF	17.39	0.6832	0.3318	0.04	0.48	51.89	43.67	0.34	0.88
+ Vis + DSDS (Ours)	18.23	0.7060	0.2867	0.02	0.28	48.67	41.31	0.35	0.69

Table 13: **"table" capture quantitative evaluation.**

References

[1] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022. 1

[2] Yuhe Jin, Dmytro Mishkin, Anastasiia Mishchuk, Jiri Matas, Pascal Fua, Kwang Moo Yi, and Eduard Trulls. Image matching across wide baselines: From paper to practice. *International Journal of Computer Vision*, 129(2):517–547, 2021. 1

[3] Ben Mildenhall, Pratul P. Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (TOG)*, 2019. 1

[4] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1