Neural 3D Reconstruction in the Wild – Supplementary Material

ACM Reference Format:
. 2022. Neural 3D Reconstruction in the Wild – Supplementary Material. In ,. ACM, New York, NY, USA, 2 pages.

1 ADDITIONAL DETAILS ON EXPERIMENTS

Training time. We provide the exact training time of the proposed method and baseline methods in Table 1. Our method is the fastest among all the baseline methods, while achieving similar or better accuricies.

Values of the evaluation thresholds. We provide the values of the error thresholds (low, mid and high) used in the evaluation in Table 2.

Vis-MVS results with different number of source views. Since our method takes all images as training data at once, to make a fair comparison, we increase the number of source views of Vis-MVS to make the depth estimation network process more images at a single inference pass. The results are presented in Table 3. It can be seen that the performance of Vis-MVS does not increase with the number of source views.

Visualization of the reconstruction error.. We visualize the reconstruction errors of our method on BG in Fig 1.

2 ADDITIONAL ABLATION STUDIES

Number of images. We provide ablation studies on the number of images used for training in Table 4. Using more images for training consistently produces better reconstruction accuricies, although the gap is not huge. We believe this is due to the fact that the cameras from Internet collections are evenly distributed among the entire scene, decreasing the number of images used for training will not significantly reduce the range of the observations, but only the density. In terms of the training time, since we are selecting the model with the best accuricies during training, the overall training time are not proportionally reduced with the number of images used.

Different voxel sizes. We provide ablation studies on different voxel sizes s in Table 5. Since increasing or decreasing s will not affect the surface-guided sampling but only the voxel-guided sampling, the results do not show any significant difference, which further signifies the effectiveness of surface-guided sampling.

Method	BG	LM	PE	PBA
NeRF-W	37.8	43.5	39.0	89.6
Vis-MVS	13.7	9.4	19.8	48.0
COLMAP	21.4	47.3	31.4	197.6
Ours 🏃	**9.7**	**7.5**	**11.5**	**14.3**
Ours	30.7	22.4	20.7	26.0

Table 1: Training time of different methods measured in hours.

Level	BG	LM	PE	PBA
Low	0.1	0.01	0.04	0.2
Medium	0.2	0.02	0.06	0.4
High	0.3	0.03	0.08	0.6
AUC-max	0.6	0.05	0.2	0.8

Table 2: Values of evaluation thresholds in meters. We report the values of low, middle and high error thresholds used in the evaluation. AUC-max represents the maximum range of the curve used to calculate the AUC.

,

source	Low			Mid			High			AUC		
views	P	R	F	P	R	F	P	R	F	P	R	F
20	**23.37**	27.04	25.07	**46.30**	42.72	44.44	**63.78**	52.98	57.88	**2.73**	2.36	2.52
40	23.02	**31.11**	**26.46**	45.72	**48.84**	**47.23**	62.35	**61.02**	**61.68**	2.67	**2.71**	**2.68**
80	20.43	28.46	23.78	42.95	46.03	44.44	61.32	57.17	59.17	2.61	2.56	2.57

Table 3: Vis-MVS results with different number of source images.

num	Low			Mid			High			AUC			Time
images	P	R	F	P	R	F	P	R	F	P	R	F	
25%	30.75	31.38	31.06	61.99	60.29	61.13	77.91	76.56	77.23	3.20	3.15	3.17	14.10
50%	30.50	33.30	31.84	62.43	62.98	62.70	78.54	79.48	79.01	3.22	3.27	3.24	14.23
75%	31.59	34.58	33.02	63.03	64.03	63.52	78.46	79.12	78.79	3.23	3.29	3.26	21.62
100%	**32.83**	**35.56**	**34.14**	**67.67**	**68.72**	**68.19**	**82.07**	**82.13**	**82.10**	**3.37**	**3.40**	**3.38**	22.42

Table 4: Effect of the number of images used for reconstruction. More images lead to better reconstruction quality. However, training also takes more time.

sfm	Low			Mid			High			AUC		
voxel	P	R	F	P	R	F	P	R	F	P	R	F
x0.5	33.11	**37.88**	35.33	65.81	68.34	67.05	80.61	81.95	81.27	3.32	**3.42**	3.37
x2.0	**34.51**	37.67	**36.02**	64.55	65.25	64.90	79.70	79.10	79.40	3.31	3.33	3.32
x1.0	32.83	35.56	34.14	**67.67**	**68.72**	**68.19**	**82.07**	**82.13**	**82.10**	**3.37**	3.40	**3.38**

Table 5: Effect of different SFM voxel sizes. Doubling or halving the SFM voxel size has little effect on the reconstruction result. The proposed surfaced-guided sampling technique generates samples centered at surface position with a predefined range, which is irrelevant to initial SFM voxel size.

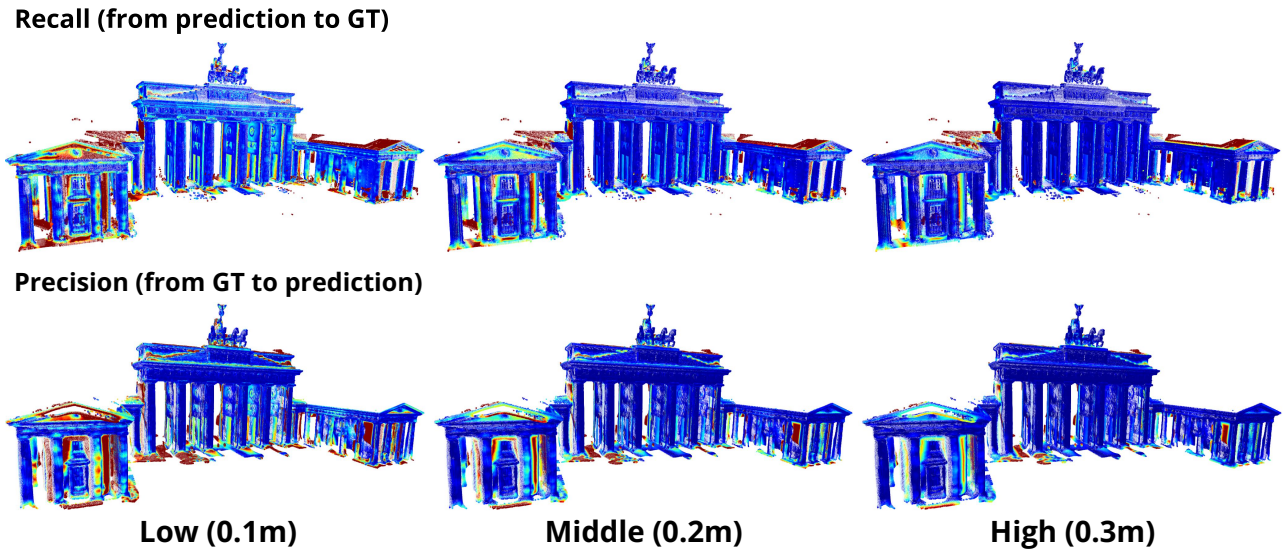

Figure 1: Visualization of the reconstruction error on BG. We visualize the reconstruction error of our method under different error thresholds. Warmer colors indicate larger errors.