Deep Permutation Equivariant Structure from Motion
-Supplementary Material-

Dror Moran[1]* Hodaya Koslowsky[1]* Yoni Kasten[1]
Haggai Maron[2] Meirav Galun[1] Ronen Basri[1]

[1]Weizmann Institute of Science [2]NVIDIA Research

Below we provide implementation details for the baseline methods and the alternative deep architectures tested in our paper. We further include additional results, including examples of reconstructions with our method.

1. Baselines

Colmap baseline. In the calibrated experiments (Table 2 in the paper), for fair comparison, we applied Colmap [9] directly to the points tracks provided by Olsson's dataset [7] and fixed the intrinsic camera parameters to those provided as ground truth.

Linear baseline. We tested Jiang et al.'s method [3] while ignoring viewing graph edges for which the number of matching points was lower than a certain threshold. We used thresholds of 30, 200, 500 matching points and report those results for which the lowest reprojection error, before bundle adjustment, was obtained.

2. Alternative deep architectures

The two right most columns in Table 5 in the paper show results of two novel deep architectures which were developed for comparison to our deep network architecture. The details are given below.

Set neural network. For a scene with m cameras, the input to this network is a set of m random feature vectors of size 12 that provide unique ids to each camera. Inspired by [8, 11], our set network is composed of three sub-networks where each sub-network is an equivariant set network. The first sub-network is applied to each feature vector and calculates a local feature for each camera. The second sub-network is applied to each such local feature. The outputs for all cameras are then averaged, producing a global scene feature vector. Finally, the camera parameters are predicted by applying the third sub-network to both the local and global feature. In summary, the set network prediction for camera i is defined as follows

$$\begin{aligned} \mathbf{z}_i &= S_1(\mathbf{v}_i) \\ \mathbf{z}_g &= \frac{1}{m}\sum_{i=1}^{m} S_2(\mathbf{z}_i) \\ P_i &= S_3(\mathbf{z}_i, \mathbf{z}_g). \end{aligned}$$

Each S_k is a fully connected network and v_i is the initial random vector of camera i.

Graph neural network. Here, the cameras are represented by the nodes of a graph, called the *viewing graph*. An edge connects a pair of nodes if the respective images share at least 30 tracks, in which case a fundamental matrix is computed. The fundamental matrices are used as edge input features, while as with the set network model, random vectors form the

*Equal contributors.

Scan	#Images	#Points	Error (pixels)					
			Before BA		After BA			
			Ours	GPSFM	**Ours**	GPSFM	PPSFM	VarPro
Alcatraz Courtyard	133	23674	**1.55**	20.34	**0.52**	**0.52**	0.57	**0.52**
Alcatraz Water Tower	172	14828	**2.18**	16.5	**0.47**	0.63	0.59	**0.47**
Alcatraz West Side Gardens	419	65072	**9.54**	1007.5	**0.76**	326.99	1.77	-
Basilica Di San Petronio	334	46035	**7.9**	1871.41	0.96	60.69	**0.63**	-
Buddah Statue	322	156356	**18.88**	919.26	2.93	96.96	**0.41**	-
Buddah Tooth Relic Temple Singapore	162	27920	**4.59**	18.53	**0.6**	0.62	0.71	**0.6**
Corridor	11	737	**0.3**	0.64	**0.26**	**0.26**	0.27	**0.26**
Ecole Superior De Guerre	35	13477	**0.75**	1.88	**0.26**	**0.26**	0.28	**0.26**
Dinosaur 319	36	319	**2.35**	4.66	1.53	**0.43**	0.47	**0.43**
Dinosaur 4983	36	4983	1.96	**1.54**	0.57	**0.42**	0.47	**0.42**
Doge Palace Venice	241	67107	**3.6**	170.93	**0.6**	3.52	0.67	-
Eglise du dome	85	84792	**1.1**	8.41	**0.24**	**0.24**	0.25	-
Drinking Fountain Somewhere In Zurich	14	5302	**0.33**	1.29	**0.28**	**0.28**	0.31	**0.28**
East Indiaman Goteborg	179	25655	**3.31**	99.38	0.99	5.11	**0.67**	-
Folke Filbyter	40	21150	8.87	**1.78**	8.58	0.82	**0.33**	277.89
Golden Statue Somewhere In Hong Kong	18	39989	**0.35**	0.81	**0.22**	**0.22**	0.24	**0.22**
Gustav Vasa	18	4249	**0.23**	1.82	**0.16**	**0.16**	0.17	**0.16**
GustavIIAdolf	57	5813	14.77	**5.91**	5.83	**0.23**	0.24	**0.23**
Model House	10	672	**0.37**	3.66	**0.34**	1.12	0.4	**0.34**
Jonas Ahlstromer	40	2021	**14.38**	28.83	4.72	**0.18**	0.2	**0.18**
Lund University Sphinx	70	32668	**3.64**	10.0	**0.34**	0.45	0.37	**0.34**
Nijo Castle Gate	19	7348	**0.71**	20.08	**0.39**	**0.39**	0.43	**0.39**
Pantheon Paris	179	29383	**1.75**	44.85	**0.49**	2.85	0.62	-
Park Gate Clermont Ferrand	34	9099	**0.61**	13.82	**0.31**	0.32	0.49	**0.31**
Plaza De Armas Santiago	240	26969	**5.1**	81.01	**0.64**	3.14	0.71	-
Porta San Donato Bologna	141	25490	**1.58**	33.36	**0.4**	0.61	3.75	**0.4**
The Pumpkin	195	69335	14.45	**8.97**	**0.38**	**0.38**	0.42	-
Skansen Kronan Gothenburg	131	28371	**1.19**	8.9	**0.41**	0.44	0.44	-
Skansen Lejonet Gothenburg	368	74423	**10.82**	69.81	2.05	7.48	**1.28**	-
Smolny Cathedral St Petersburg	131	51115	**1.66**	83.78	**0.46**	**0.46**	0.5	-
Some Cathedral In Barcelona	177	30367	**3.67**	14.77	**0.51**	**0.51**	0.54	-
Sri Mariamman Singapore	222	56220	**7.06**	39.89	**0.61**	0.78	0.85	-
Sri Thendayuthapani Singapore	98	88849	**2.12**	13.25	**0.31**	0.56	0.33	-
Sri Veeramakaliamman Singapore	157	130013	**6.47**	99.99	**0.52**	1.78	0.66	-
Thian Hook Keng Temple Singapore	138	34288	**7.59**	26.78	**0.54**	0.55	0.66	**0.54**
King's College University Of Toronto	77	7087	**2.27**	22.89	0.78	2.35	0.26	**0.24**
Tsar Nikolai I	98	37857	**6.04**	13.21	2.43	0.33	0.31	**0.29**
Urban II	96	22284	**16.91**	87.25	6.84	**0.27**	0.31	3.61

Table 1: Single scene experiments in the uncalibrated setup. The table shows mean reprojection errors obtained with our method before and after BA, compared to GPSFM [5], PPSFM [6] and VarPro [2]. (*Smaller is better.*) Our comparison to VarPro is partial, since in a number of experiments it exceeded either memory or runtime limitations.

node input features. We use a message-passing scheme [1] and global feature as described for the set network model. Each message-passing layer is of the following form

$$\mathbf{z}_i^l = \frac{1}{|N_i|} \sum_{j \in N_i} \mathrm{mlp}_l(\mathbf{z}_i^{l-1}, \mathbf{z}_j^{l-1}, F_{ij})$$

where $\mathbf{z}_i^l$ is the local feature of node i in layer l, N_i are the neighbors of node i and F_{ij} is the fundamental matrix measured between cameras i and j.

Both the set and the graph models predict camera parameters, while the 3D points are treated as free variables. In both cases we minimize the reprojection loss defined in equation (3) in the paper.

3. Results

Single scene recovery. In the single scene recovery mode given a single track tensor representing point correspondences across images of some scene we attempt to minimize the reprojection loss, where the network is used to parameterize the loss. Tables 1 and 2 show results of our method before and after bundle adjustment in the uncalibrated and calibrated settings. We compare our results before bundle adjustment only to global methods since sequential methods apply bundle adjustment in each iteration. Notably, already before bundle adjustment our method often achieves sub-pixel accuracies, significantly surpassing GPSFM in the uncalibrated setting and GESFM and Linear in the calibrated setting. Figures 1-3 show 3D reconstructions and camera parameter recovery in the calibrated setting. In addition, a failure case is shown in Figure 4. Figure 5 shows the evolution of structure and camera parameters during optimization.

Scan	#Images	#Points	Before BA									After BA										
			$\mathbf{t}_{\text{error}}$			R_{error}			Reprojection Err.			$\mathbf{t}_{\text{error}}$				R_{error}				Reprojection Err.			
			Ours	GESFM	Linear	**Ours**	GESFM	Linear	**Ours**	GESFM	Linear	**Ours**	GESFM	Linear	Colmap	**Ours**	GESFM	Linear	Colmap	**Ours**	GESFM	Linear	Colmap
Alcatraz Courtyard	133	23674	**0.16**	0.767	0.378	**0.619**	1.851	0.729	**1.64**	66.5	16.58	0.015	0.259	**0.014**	**0.014**	0.049	0.533	**0.042**	0.043	**0.81**	4.67	1.27	**0.81**
Alcatraz Water Tower	172	14828	**0.518**	8.332	1.643	**0.933**	1.136	1.525	**2.13**	131.81	56.26	0.116	9.147	1.643	**0.115**	0.23	9.997	1.525	**0.228**	**0.55**	25.93	73.72	**0.55**
Buddah Tooth Relic Temple Singapore	162	27920	**0.233**	2.124	1.325	**1.03**	2.95	2.058	**2.06**	89.94	47.5	**0.014**	1.429	0.125	0.015	**0.081**	4.709	0.551	0.083	**0.85**	13.22	2.66	**0.85**
Doge Palace Venice	241	67107	**0.342**	1.688	-	**1.163**	2.75	-	**3.62**	123.53	-	0.029	1.608	-	**0.012**	0.211	5.317	-	**0.031**	1.0	22.32	-	**0.98**
Door Land	12	17650	**0.006**	(1.603)	0.226	**0.024**	(2.041)	1.148	**0.32**	(227.0)	20.89	**0.001**	(0.973)	**0.001**	**0.001**	0.006	(7.552)	**0.005**	**0.005**	**0.3**	(9.21)	**0.3**	**0.3**
Drinking Fountain Somewhere In Zurich	14	5302	**0.004**	(0.016)	0.024	**0.031**	(0.054)	0.077	**0.33**	(0.94)	0.58	**0.002**	(0.002)	**0.002**	**0.002**	**0.007**	(0.01)	**0.007**	**0.007**	0.31	(0.27)	0.31	0.31
East Indiaman Goteborg	179	25655	**0.621**	2.783	(2.235)	3.814	11.129	(3.284)	**4.13**	170.63	(94.46)	0.509	3.099	(2.235)	**0.065**	3.117	12.396	(3.284)	**0.251**	1.85	32.37	(312.9)	**0.89**
Ecole Superior De Guerre	35	13477	0.081	(0.006)	0.048	0.318	(0.057)	0.182	0.72	(0.35)	1.48	0.005	(0.002)	0.005	0.005	**0.024**	(0.035)	**0.024**	**0.024**	0.34	(0.14)	0.34	0.34
Eglise du dome	85	84792	0.205	(1.958)	**0.128**	**0.808**	(2.851)	0.903	**0.91**	(90.83)	26.4	**0.01**	(1.425)	0.046	**0.01**	0.037	(3.631)	0.162	**0.036**	**0.27**	(6.21)	0.76	**0.27**
Folke Filbyter	40	21150	0.125	(0.003)	0.021	74.596	(0.332)	1.94	10.37	(5.74)	72.06	0.118	(0.0)	0.123	**0.0**	70.157	(0.148)	4.484	**0.036**	4.29	(0.41)	6.06	**0.29**
Fort Channing Gate Singapore	27	23627	0.093	**0.092**	0.139	**0.207**	0.295	0.659	**0.52**	2.57	22.69	**0.008**	**0.008**	0.013	**0.008**	**0.02**	**0.02**	0.029	**0.02**	**0.25**	**0.25**	0.45	**0.25**
Golden Statue Somewhere In Hong Kong	18	39989	**0.073**	0.118	1.153	**0.292**	0.669	8.264	**0.4**	4.98	73.7	**0.004**	**0.004**	**0.004**	**0.004**	0.031	0.03	**0.022**	0.031	**0.27**	**0.27**	0.3	**0.27**
Gustav Vasa	18	4249	1.085	(0.079)	0.266	34.181	(0.841)	1.658	**3.52**	(5.21)	11.99	1.145	(0.101)	**0.099**	0.1	32.266	(0.751)	0.839	0.841	3.15	(0.31)	0.48	0.48
GustavIIAdolf	57	5813	9.714	**0.134**	0.333	67.784	**0.435**	1.398	13.91	**6.49**	31.08	8.524	**0.004**	**0.004**	**0.004**	58.458	**0.021**	**0.021**	**0.021**	11.49	**0.26**	**0.26**	**0.26**
Jonas Ahlstromer	40	2021	10.888	(0.35)	0.895	50.19	(1.994)	10.154	**10.82**	(36.48)	236.41	10.451	(0.01)	1.259	0.011	47.117	(0.082)	5.391	**0.036**	8.41	(0.69)	4.69	**0.22**
King's College University Of Toronto	77	7087	0.235	(0.152)	(1.781)	0.989	(0.645)	(1.07)	**0.9**	(11.87)	(27.29)	0.017	(0.005)	(1.877)	0.017	0.085	(0.059)	(4.624)	0.084	**0.34**	(0.35)	(7.12)	**0.34**
Lund University Sphinx	70	32668	4.585	**0.228**	1.199	19.522	**0.738**	3.476	**4.78**	7.19	60.64	2.191	0.016	1.512	**0.009**	8.752	0.058	5.452	**0.033**	1.36	0.4	4.58	**0.39**
Nijo Castle Gate	19	7348	0.286	**0.141**	0.348	1.495	**0.399**	2.097	**1.7**	11.18	154.96	0.012	**0.011**	0.19	**0.011**	0.069	**0.064**	0.744	**0.064**	**0.73**	**0.73**	4.84	**0.73**
Pantheon Paris	179	29383	**0.05**	0.867	1.275	**0.192**	3.766	2.655	**1.47**	79.24	39.69	**0.005**	0.595	0.011	-	**0.04**	3.208	0.072	-	**0.49**	9.71	0.82	-
Park Gate Clermont Ferrand	34	9099	0.125	**0.083**	0.1	0.391	**0.203**	0.296	**0.57**	1.71	10.5	**0.022**	**0.022**	**0.022**	**0.022**	**0.049**	**0.049**	**0.049**	**0.049**	**0.35**	**0.35**	**0.35**	**0.35**
Plaza De Armas Santiago	240	26969	2.944	**2.45**	-	6.782	**6.291**	-	**7.4**	146.56	-	1.383	2.244	-	**0.048**	2.556	6.344	-	**0.122**	4.9	15.61	-	**1.13**
Porta San Donato Bologna	141	25490	**0.388**	0.949	1.588	2.153	**1.013**	1.381	**2.28**	29.5	46.12	**0.046**	0.169	0.067	0.047	**0.095**	0.513	0.149	0.099	**0.75**	3.23	1.16	**0.75**
Round Church Cambridge	92	84643	1.003	0.486	**0.217**	2.451	1.021	**0.634**	**2.66**	19.04	9.6	0.582	0.493	**0.012**	**0.012**	1.107	1.851	**0.033**	0.035	1.54	2.03	0.41	**0.39**
Skansen Kronan Gothenburg	131	28371	0.226	**0.223**	(0.234)	0.736	**0.549**	(0.679)	**1.24**	8.82	(18.49)	0.008	0.008	(0.007)	0.008	0.026	0.025	(0.02)	0.025	**0.67**	**0.67**	(0.69)	**0.67**
Smolny Cathedral St Petersburg	131	51115	**0.051**	0.209	-	0.554	**0.493**	-	**1.66**	19.01	-	**0.006**	0.007	-	**0.006**	0.033	**0.028**	-	0.029	**0.81**	1.0	-	**0.81**
Some Cathedral In Barcelona	177	30367	**0.315**	1.776	1.261	**0.88**	1.519	3.126	**2.87**	47.12	66.97	0.011	0.013	0.024	**0.01**	0.026	0.031	0.057	**0.025**	**0.89**	1.09	2.09	**0.89**
Sri Mariamman Singapore	222	56220	**0.683**	1.758	0.721	2.302	**1.433**	1.615	**4.13**	52.13	37.16	**0.023**	0.614	0.025	**0.023**	**0.077**	2.158	0.083	0.078	0.91	7.4	1.17	**0.89**
Sri Thendayuthapani Singapore	98	88849	3.812	(0.285)	0.375	46.269	(1.561)	1.581	23.37	(15.93)	19.57	2.87	(0.053)	**0.034**	**0.034**	44.17	(0.329)	**0.138**	**0.138**	8.44	(0.56)	0.72	0.67
Sri Veeramakaliamman Singapore	157	130013	0.597	(1.966)	**0.273**	2.559	(1.807)	**0.519**	**3.47**	(205.96)	18.08	0.04	(1.388)	0.095	**0.038**	0.175	(3.41)	0.288	**0.169**	0.73	(34.72)	2.2	**0.71**
Statue Of Liberty	134	49250	20.012	(4.55)	**3.031**	46.887	(3.449)	**3.357**	**26.16**	(1031.8)	133.81	4.122	(4.782)	28.049	**0.099**	9.091	(8.281)	2.945	**0.213**	6.97	(52.05)	5.08	**1.25**
The Pumpkin	196	69341	14.89	**0.513**	(1.656)	94.672	**2.036**	(4.215)	33.41	**9.71**	(122.54)	14.952	**0.022**	(14.862)	**0.022**	98.862	0.092	(3.123)	**0.091**	24.85	**0.57**	(24.19)	**0.57**
Thian Hook Keng Temple Singapore	138	34288	**0.082**	0.519	0.404	**0.832**	2.751	3.047	**2.75**	53.79	62.7	**0.008**	0.024	0.043	**0.008**	**0.081**	0.245	0.424	0.084	1.13	3.32	4.92	**1.12**
Tsar Nikolai I	98	37857	9.467	**0.219**	0.261	48.499	**0.475**	1.437	9.79	**5.19**	32.86	7.836	**0.005**	**0.005**	**0.005**	36.28	**0.018**	**0.018**	**0.018**	6.53	**0.33**	**0.33**	**0.33**
Urban II	96	22284	9.467	**0.774**	2.044	47.49	**2.077**	8.951	**9.38**	31.71	176.19	9.586	0.036	3.038	**0.021**	48.214	0.175	16.348	**0.107**	6.92	0.72	17.61	**0.38**
Vercingetorix	69	10754	8.788	**1.158**	2.786	69.328	**2.203**	2.365	**5.08**	15.87	65.57	3.104	0.3	1.564	**0.011**	17.706	1.431	7.138	**0.048**	1.5	0.54	2.93	**0.23**
Yueh Hai Ching Temple Singapore	43	13774	**0.098**	(0.642)	0.303	**0.72**	(1.813)	1.92	**0.94**	(27.32)	45.19	**0.014**	(0.023)	0.059	**0.014**	**0.043**	(0.075)	0.26	**0.043**	**0.65**	(1.64)	2.06	**0.65**

Table 2: Single scene experiments in the calibrated setup. The table shows mean camera location error (denoted $\mathbf{t}_{\text{error}}$) in meters, mean orientation error (denoted R_{error}) in degrees, and mean reprojection error in pixels obtained with our method before and after BA, compared to GESFM [4], Linear [3], and Colmap [10]. (*Smaller is better.*) In parenthesis experiments in which at least 10% of the cameras are removed.

Scan	#Images	#Points	$\mathbf{t}_{\text{error}}$			R_{error}			Reprojection Error		
			Ours No BA	**Ours**	Colmap	**Ours No BA**	**Ours**	Colmap	**Ours No BA**	**Ours**	Colmap
Folke Filbyter	40	21150	0.093	0.037	**0.0**	54.025	20.51	**0.036**	9.78	3.87	**0.29**
Gustav Vasa	18	4249	0.193	**0.099**	0.1	2.964	**0.839**	0.841	0.62	**0.48**	**0.48**
GustavIIAdolf	57	5813	0.014	**0.004**	**0.004**	0.068	**0.021**	**0.021**	0.29	**0.26**	**0.26**
Jonas Ahlstromer	40	2021	0.018	**0.011**	**0.011**	0.051	0.037	**0.036**	0.24	**0.22**	**0.22**
Plaza De Armas Santiago	240	26969	**0.044**	0.048	0.048	**0.089**	0.121	0.122	1.3	**1.13**	**1.13**
Sri Thendayuthapani Singapore	98	88849	0.057	**0.034**	**0.034**	0.222	0.139	**0.138**	0.77	**0.67**	**0.67**
Statue Of Liberty	134	49250	8.558	1.877	**0.099**	13.262	3.02	**0.213**	6.86	1.76	**1.25**
The Pumpkin	196	69341	0.17	**0.022**	**0.022**	0.851	0.092	**0.091**	0.7	**0.57**	**0.57**
Tsar Nikolai I	98	37857	0.024	**0.005**	**0.005**	0.092	**0.018**	**0.018**	0.38	**0.33**	**0.33**
Urban II	96	22284	0.074	**0.021**	**0.021**	0.327	**0.107**	**0.107**	0.53	**0.38**	**0.38**

Table 3: Single scene results using sequential optimization in the calibrated setup. The table show results before and after bundle adjustment compared to Colmap. The table shows mean camera location error (denoted $\mathbf{t}_{\text{error}}$) in meters, mean orientation error (denoted R_{error}) in degrees, and mean reprojection error in pixels. (*Smaller is better.*)

Sequential optimization. In some experiments, as can be seen in Table 2, our single scene recovery procedure failed to produce accurate reconstruction. In these cases (we declared failure if the reprojection error exceeded 2 pixels) we applied instead optimization at a sequential schedule. For this schedule we ordered the images greedily by the number of point tracks they share with the images that precede them in this order. Using this order, we first ran 500 optimization epochs with just the first 2 images. Then, after each 500 more epochs we add to this subset the next image in the order. As can be seen in Table 3, this optimization schedule improved the reprojection error for all the failed datasets, yielding in most cases comparable accuracies to those obtained with Colmap.

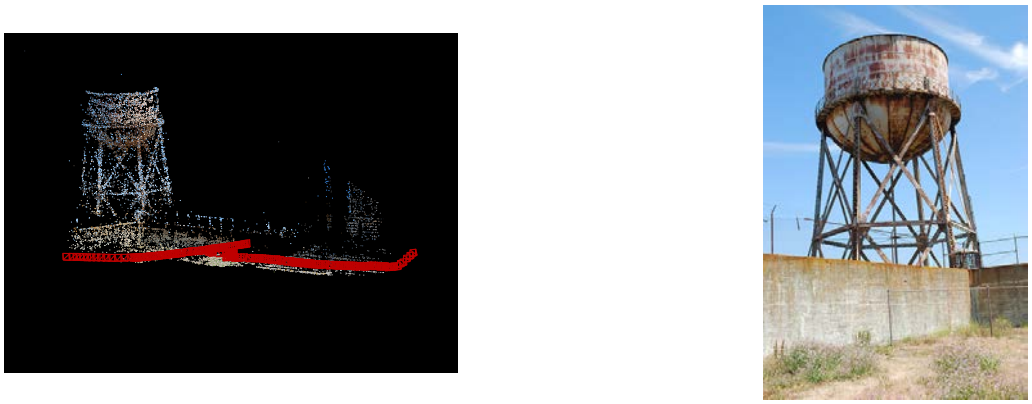
Scan	#Images	#Points	Time (seconds)			
			Inference	Fine tuning	BA	Colmap
Alcatraz Courtyard	133	23674	0.007	199.125	43.512	286.0
Alcatraz Water Tower	172	14828	0.007	110.847	26.44	130.0
Drinking Fountain Somewhere In Zurich	14	5302	0.007	20.302	2.925	16.0
Nijo Castle Gate	19	7348	0.008	24.493	4.308	21.0
Porta San Donato Bologna	141	25490	0.007	194.651	45.416	170.0
Round Church Cambridge	92	84643	0.014	360.97	90.092	229.0
Smolny Cathedral St Petersburg	131	51115	0.004	534.528	101.456	516.0
Some Cathedral In Barcelona	177	30367	0.007	208.542	55.424	451.0
Sri Veeramakaliamman Singapore	157	130013	0.319	291.727	242.888	583.0
Yueh Hai Ching Temple Singapore	43	13774	0.004	45.458	10.539	106.0

Scan	#Images	#Points	Time (seconds)			
			Inference	Fine tuning	BA	GPSFM
Alcatraz Water Tower	172	14828	0.055	89.646	68.939	137.057
Dino 319	36	319	0.004	8.151	0.475	3.253
Dino 4983	36	4983	0.122	7.66	2.21	4.994
Dome	85	84792	0.203	160.896	76.867	105.837
Drinking Fountain	14	5302	0.01	18.197	3.016	3.348
Gustav Vasa	18	4249	0.007	14.435	2.766	3.449
Nijo	19	7348	0.013	34.397	3.121	6.37
Skansen Kronan	131	28371	0.141	197.531	63.853	93.831
Some Cathedral In Barcelona	177	30367	0.133	185.984	47.597	110.485
Sri Veeramakaliamman Singapore	157	130013	0.314	473.294	195.374	301.713

Table 4: Execution times for our trained model. The table shows execution times in seconds in the calibrated (left) and uncalibrated (right) settings.

Learning from multiple scenes. Figures 6 and 7 show reconstruction results using our model before and after bundle adjustment in 3 scenarios: (i) inference using our trained model (ii) inference followed by fine tuning and (iii) short run of optimization. Table 4 shows execution times for our trained model. We note that using inference only yields a good initialization for bundle adjustment in a small fraction of a second. Using fine tuning yields more accurate results (See Table 2 in the paper) with execution times similar to Colmap. The short optimization generally yields less accurate results with execution times similar to fine tuning, emphasizing the importance of the trained model.

(a) Alcatraz Courtyard

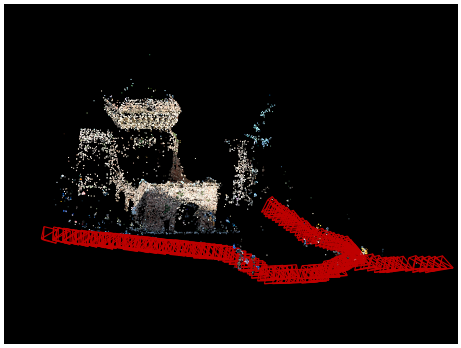


(b) Alcatraz Water Tower

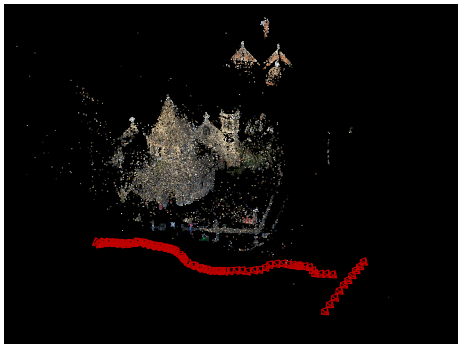
(c) Drinking Fountain Somewhere In Zurich

(d) Nijo Castle Gate

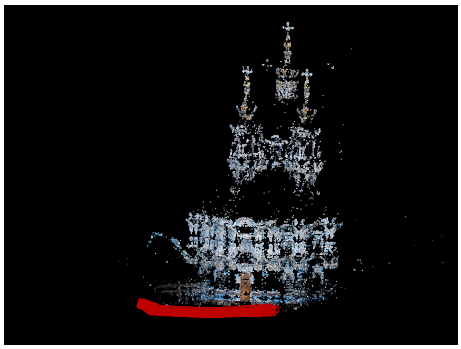
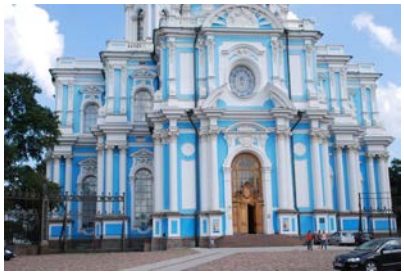
Figure 1: Single scene 3D reconstructions and recovery of camera parameters with our method. Each pair shows on the left the triangulated point cloud and the recovered camera locations and orientations (in red) and on the right one of the input images.



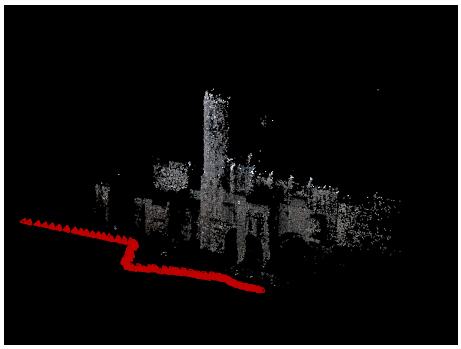

(a) Porta San Donato Bologna




(b) Round Church Cambridge

(c) Smolny Cathedral St Petersburg




(d) Some Cathedral In Barcelona

Figure 2: Single scene 3D reconstructions and recovery of camera parameters with our method. Each pair shows on the left the triangulated point cloud and the recovered camera locations and orientations (in red) and on the right one of the input images.

(a) Yueh Hai Ching Temple Singapore

Figure 3: Single scene 3D reconstructions and recovery of camera parameters with our method. Each pair shows on the left the triangulated point cloud and the recovered camera locations and orientations (in red) and on the right one of the input images.

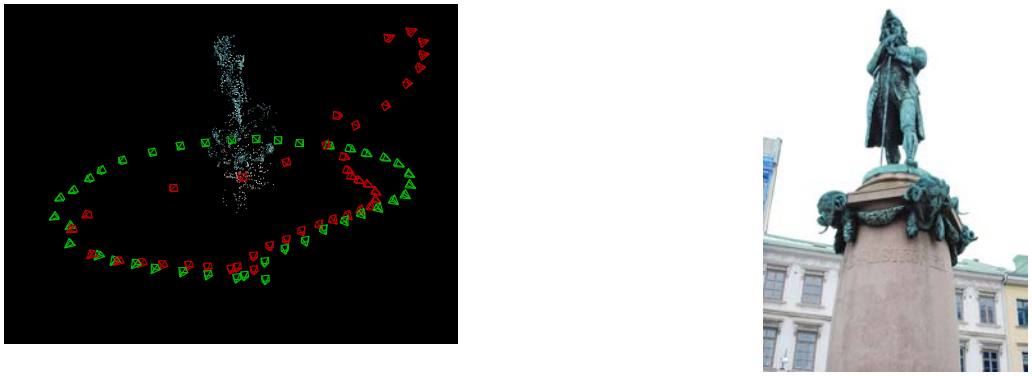


Figure 4: A failure case. Single scene 3D reconstruction and recovery of camera parameters with our method applied to Jonas Ahlstromer (reprojection error 8.41 pixels). The left image shows the triangulated point cloud, the recovered camera locations and orientations (in red) and the ground truth camera locations and orientations (in green). The right image is one of the input images.

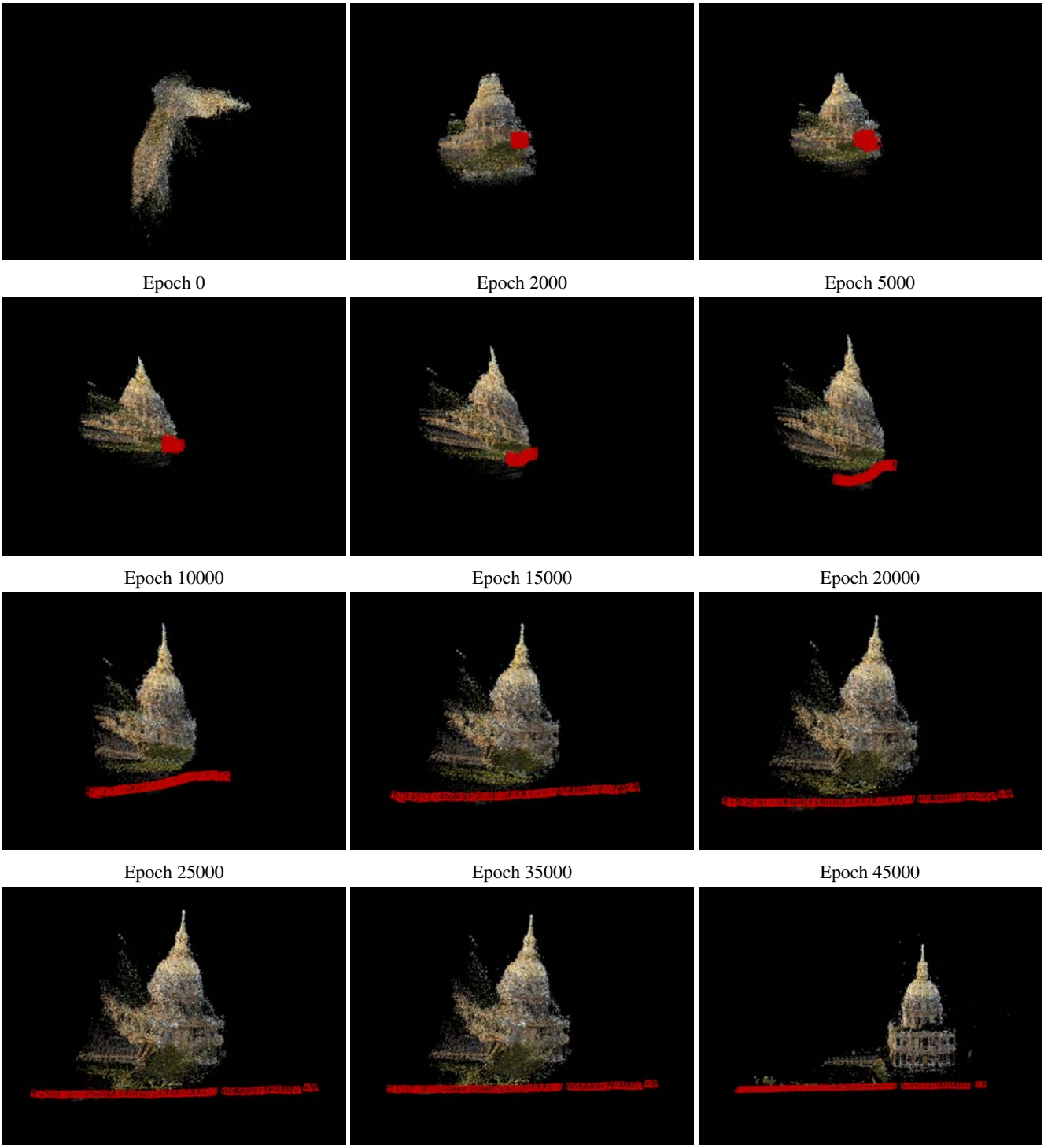
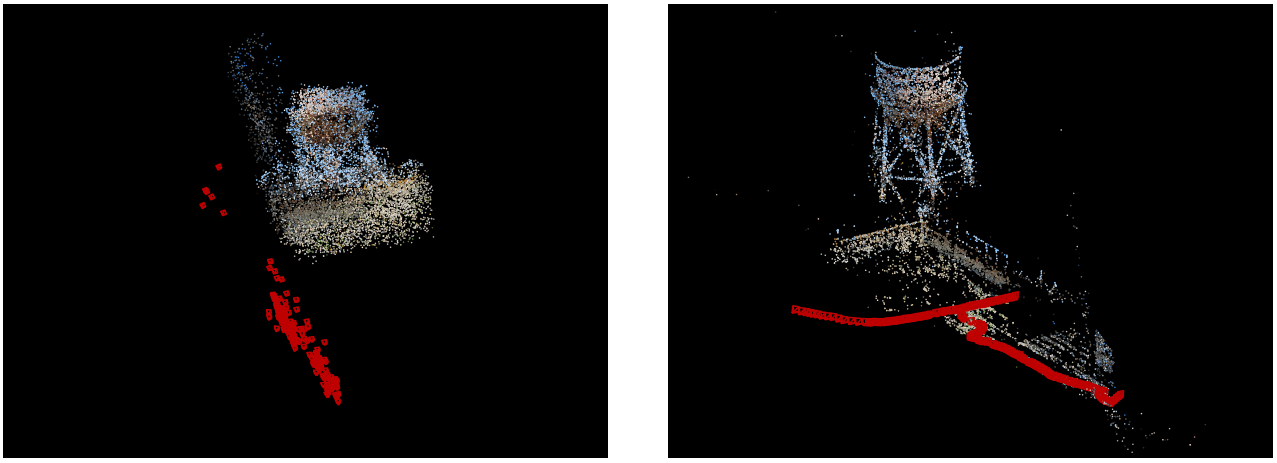
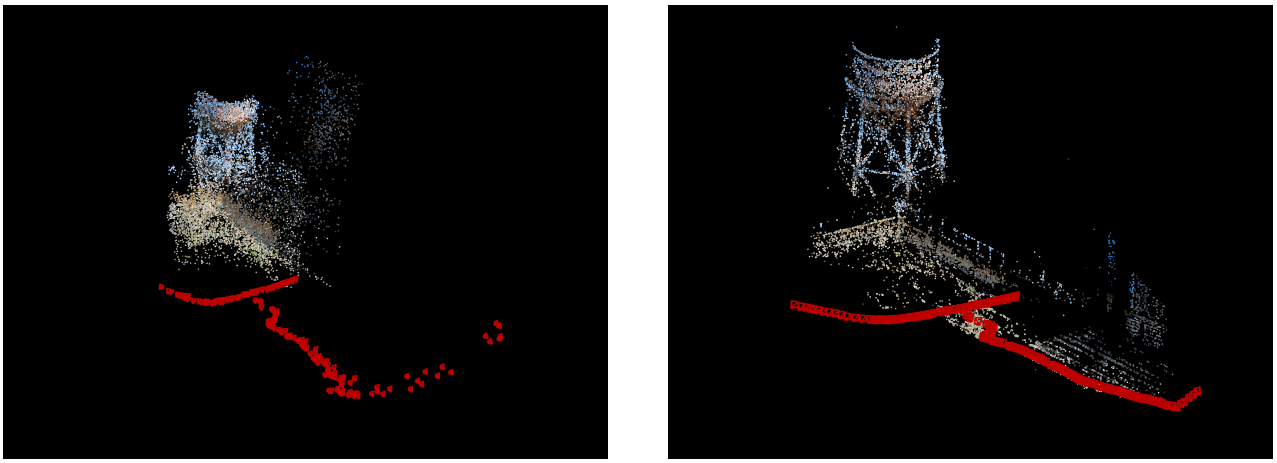


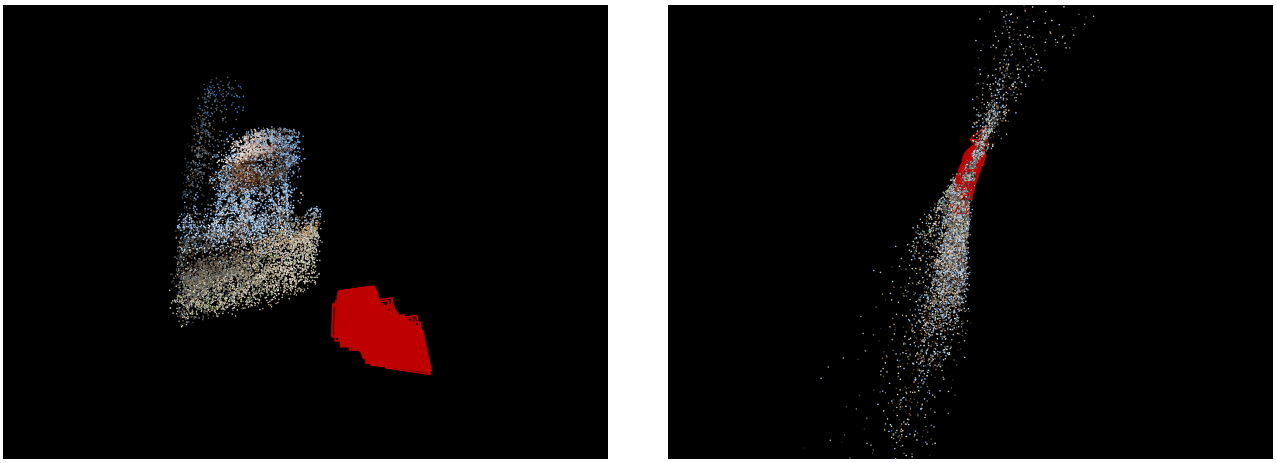
Figure 5: 3D structure and camera parameter evolution during the optimization of the network.



(a) Inference before (left) and after (right) BA

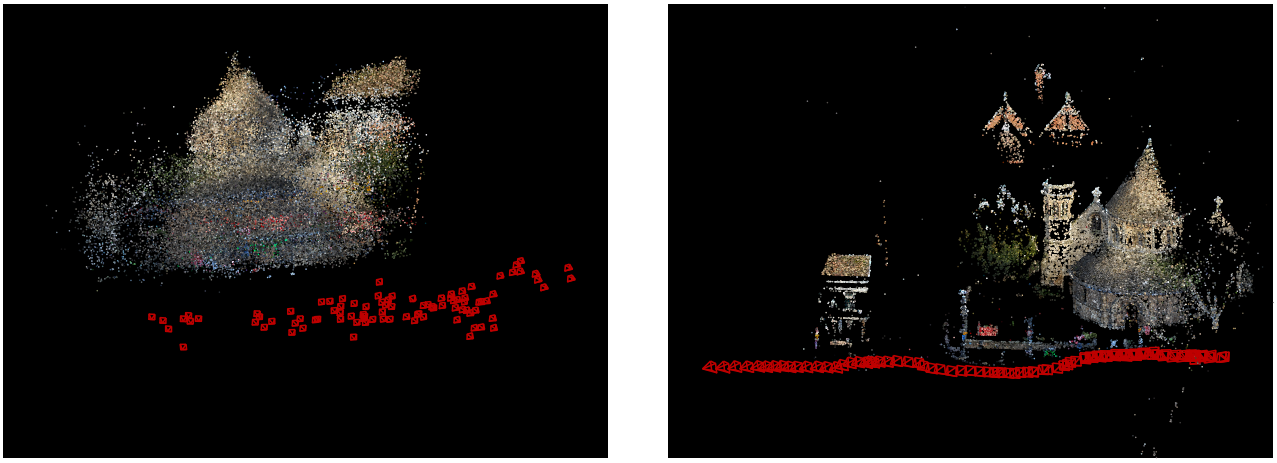


(b) Inference + fine tuning before (left) and after (right) BA

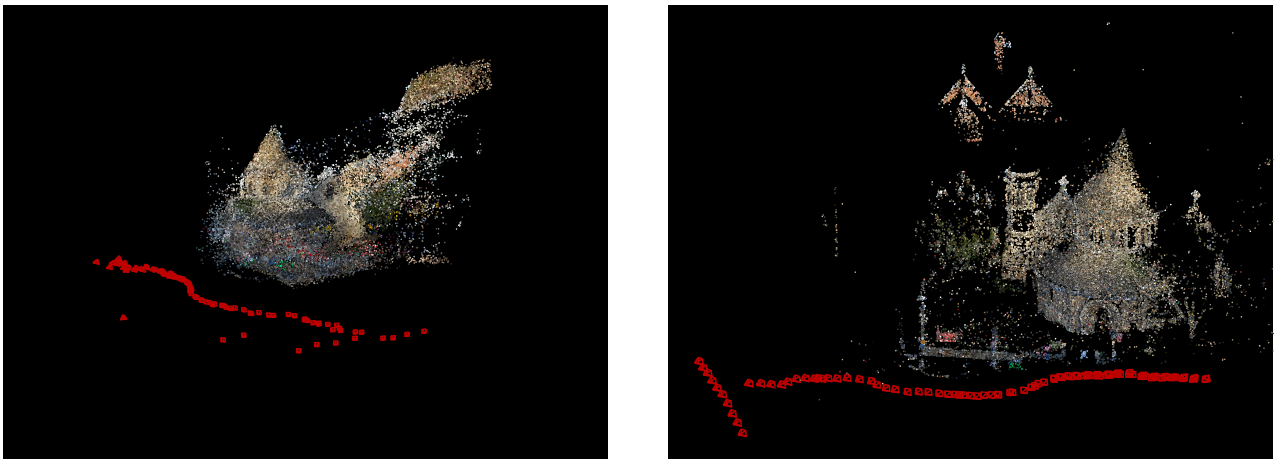


(c) Short optimization before (left) and after (right) BA

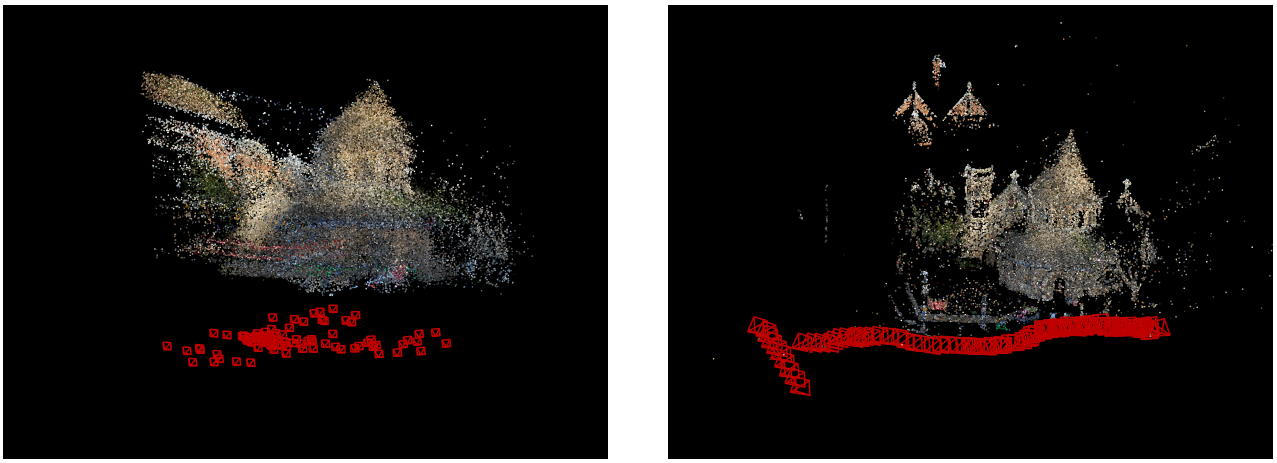
Figure 6: Alcatraz Water Tower. Reconstruction with our trained model. The figure shows results of inference (top row) and inference followed by fine tuning (middle row) before BA (left) and after BA (right). The bottom row shows the result of the short optimization strategy (starting with a random initialization). Each panel shows the recovered cameras positions (in red) and the recovered 3D points, corresponding to the point tracks. It can be seen that in this case accurate reconstruction can be obtained either by pure inference or inference followed by fine tuning (+ BA). In contrast, short optimization failed to accurately recover camera positions, leading to failure of the BA.



(a) Inference before (left) and after (right) BA



(b) Inference + fine tuning before (left) and after (right) BA



(c) Short optimization before (left) and after (right) BA

Figure 7: Round Church Cambridge. Reconstruction with our trained model. The figure shows results of inference (top row) and fine tuning (middle row) before BA (left) and after BA (right). The bottom row shows the result of the short optimization (starting with a random initialization). Each panel shows the recovered cameras positions (in red) and 3D points. Here fine tuning + BA yielded the most accurate reconstruction.

References

[1] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pages 1263–1272. PMLR, 2017. 2

[2] Je Hyeong Hong, Christopher Zach, Andrew Fitzgibbon, and Roberto Cipolla. Projective bundle adjustment from arbitrary initialization using the variable projection method. In *European Conference on Computer Vision*, pages 477–493. Springer, 2016. 2

[3] Nianjuan Jiang, Zhaopeng Cui, and Ping Tan. A global linear method for camera pose registration. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 481–488, 2013. 1, 3

[4] Yoni Kasten, Amnon Geifman, Meirav Galun, and Ronen Basri. Algebraic characterization of essential matrices and their averaging in multiview settings. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5895–5903, 2019. 3

[5] Yoni Kasten, Amnon Geifman, Meirav Galun, and Ronen Basri. Gpsfm: Global projective sfm using algebraic constraints on multiview fundamental matrices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3264–3272, 2019. 2

[6] Ludovic Magerand and Alessio Del Bue. Practical projective structure from motion (p2sfm). In *Proceedings of the IEEE International Conference on Computer Vision*, pages 39–47, 2017. 2

[7] Carl Olsson and Olof Enqvist. Stable structure from motion for unordered image collections. In *Scandinavian Conf. on Image Analysis*, pages 524–535. Springer, 2011. 1

[8] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 1

[9] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 1

[10] Johannes L. Schönberger. Colmap code. https://colmap.github.io/. 3

[11] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan R Salakhutdinov, and Alexander J Smola. Deep sets. In *Advances in neural information processing systems*, pages 3391–3401, 2017. 1