Learning Structure-from-Motion with Graph Attention Networks

Supplementary Material

In this supplementary material, we provide additional implementation details (Section A), dataset statistics (Section B), and some elaboration on alignment between estimated and ground truth camera poses carried out in certain cases (Section C). Finally, additional results are presented in Section D, including visualizations of the reconstructions quantitatively evaluated in the main paper.

A. Implementation Details

A.1. Artificial Outlier Injection Details

Here we explain the scheme we use when choosing which keypoints to corrupt with artificial outliers:

1. Define a desired outlier rate η, which we set to 10% in our experiments.
2. Mark all keypoints in views with ≤ 8 scene points visible as inliers, and do the same for all projections of scene points visible in ≤ 2 views. Let N denote the total number of keypoints, and $n_{fixed\_inliers}$ denote the number of keypoints marked as inliers here.
3. From the $N - n_{fixed\_inliers}$ remaining keypoints, take a random sample of candidates for outlier injection. In order not to violate the view and scene point lower bounds, some of these may not be accepted as candidates, so we add some margin on the number of keypoints selected. If selecting n_{sel_target} out of the remaining $N - n_{fixed\_inliers}$ candidates, i.e. a fraction of $\nu = \frac{n_{sel\_target}}{N - n_{fixed\_inliers}}$, would result in the desired outlier rate η, we instead sample a fraction according to the the harmonic average of ν and 1, i.e. $1/(0.5 \cdot \frac{1}{\nu} + 0.5 \cdot \frac{1}{1})$..
4. We now determine whether the lower bounds would be violated for any views or scene points, by marking the candidates as outliers. If so, all projections corresponding to any of those views or scene points are instead fixed to be inliers.
5. If the remaining number of outlier candidates are still enough to meet the desired outlier rate, a random sample of the candidates is selected, to achieve exactly the desired rate η. If the number of candidates are not enough, we repeat steps 3.-4. until the condition is met.

A.2. Training and Learning Rate Schedule

We train our model for 40k epochs and perform validation every 250 epochs. Since we are using an attention-based model, we warm up the learning rate for 2500 iterations, linearly from 0 to 1e-4. Then we apply an exponential learning rate decay corresponding to a factor of 10 every 250k iterations. In case of fine-tuning, the learning rate is kept constant at 1e-4 through all 1000 iterations. By the same approach as [31], we randomly sample partial scenes during training, with subsequences of 10-20 views.

A.3. Computational Resources

We carry out our experiments on a compute cluster with Nvidia A40 GPUs.

A.4. Post-Processing

Post-processing of the network output is done in the same manner as [31]. That is, either a cheap DLT triangulation is carried out, often improving the quality of the reconstruction, or bundle adjustment (BA) is carried out. BA, if applied, is carried out with Huber loss with threshold 0.1, and in two rounds interleaved with triangulation, according to [31]. This also tends to improve the final results a bit, if global convergence is not achieved during the first round. Each round of BA is limited to 100 iterations.

B. Dataset Statistics

In Tables 4 and 5 we report the number of views and scene points for each of the Euclidean and projective scenes, respectively.

C. Camera Pose Alignment

For both Euclidean and projective reconstruction, there is always a respective inherent ambiguity regarding the choice of coordinate frame. The loss function that we use is invariant to the choice of coordinate frame, but we also carry out quantitative and qualitative evaluations of the reconstructed scene points and estimated camera poses in the Euclidean setting. In order to do so, however, one first needs to determine what coordinate frame to use. To this end, in the same manner as [31], we fit a similarity transformation between estimated and ground truth camera poses, by which the two may be aligned. It is, however, worth noting that this is not done with a lot of care / robustness, and the alignment may be quite arbitrary in case the pose estimates are inaccurate.

D. Additional Results

D.1. Euclidean Reconstruction of Novel Scenes

In addition to the Euclidean reconstruction results presented in Section 4.2, which focused on network inference followed by bundle adjustment, in this section we also consider fine-tuning of the network parameters on the novel scenes, before carrying out bundle adjustment. Fine-tuning is initialized with the model parameters at the epoch of

	#Views	#Scene Points
Alcatraz Courtyard	133	23674
Alcatraz Water Tower	172	14828
Buddah Tooth Relic Temple Singapore	162	27920
Doge Palace Venice	241	67107
Door Lund	12	17650
Drinking Fountain Somewhere in Zurich	14	5302
East Indiaman Goteborg	179	25655
Ecole Superior De Guerre	35	13477
Eglise Du Dome	85	84792
Folke Filbyter	40	21150
Fort Channing Gate Singapore	27	23627
Golden Statue Somewhere in Hong Kong	18	39989
Gustav II Adolf	57	5813
Gustav Vasa	18	4249
Jonas Ahlstromer	40	2021
King´s College University of Toronto	77	7087
Lund University Sphinx	70	32668
Nijo Castle Gate	19	7348
Pantheon Paris	179	29383
Park Gate Clermont Ferrand	34	9099
Plaza De Armas Santiago	240	26969
Porta San Donato Bologna	141	25490
Round Church Cambridge	92	84643
Skansen Kronan Gothenburg	131	28371
Smolny Cathedral St Petersburg	131	51115
Some Cathedral in Barcelona	177	30367
Sri Mariamman Singapore	222	56220
Sri Thendayuthapani Singapore	98	88849
Sri Veeramakaliamman Singapore	157	130013
Statue Of Liberty	134	49250
The Pumpkin	196	69341
Thian Hook Keng Temple Singapore	138	34288
Tsar Nikolai I	98	37857
Urban II	96	22284
Vercingetorix	69	10754
Yueh Hai Ching Temple Singapore	43	13774

Table 4. Number of views and scene points for the Euclidean scenes.

	#Views	#Scene Points
Alcatraz Courtyard	133	23674
Alcatraz Water Tower	172	14828
Alcatraz West Side Gardens	419	65072
Basilica Di San Petronio	334	46035
Buddah Statue	322	156356
Buddah Tooth Relic Temple Singapore	162	27920
Corridor	11	737
Dinosaur 319	36	319
Dinosaur 4983	36	4983
Doge Palace Venice	241	67107
Drinking Fountain Somewhere in Zurich	14	5302
East Indiaman Goteborg	179	25655
Ecole Superior De Guerre	35	13477
Eglise Du Dome	85	84792
Folke Filbyter	40	21150
Golden Statue Somewhere in Hong Kong	18	39989
Gustav II Adolf	57	5813
Gustav Vasa	18	4249
Jonas Ahlstromer	40	2021
King´s College University of Toronto	77	7087
Lund University Sphinx	70	32668
Model House	10	672
Nijo Castle Gate	19	7348
Pantheon Paris	179	29383
Park Gate Clermont Ferrand	34	9099
Plaza De Armas Santiago	240	26969
Porta San Donato Bologna	141	25490
Skansen Kronan Gothenburg	131	28371
Skansen Lejonet Gothenburg	368	74423
Smolny Cathedral St Petersburg	131	51115
Some Cathedral in Barcelona	177	30367
Sri Mariamman Singapore	222	56220
Sri Thendayuthapani Singapore	98	88849
Sri Veeramakaliamman Singapore	157	130013
The Pumpkin	195	69335
Thian Hook Keng Temple Singapore	138	34288
Tsar Nikolai I	98	37857
Urban II	96	22284

Table 5. Number of views and scene points for the projective scenes.

minimal average reprojection error on the validation set. It should be stressed that this optimization is very costly, and not something that we advocate. We merely present these experiments to provide a complete comparison with DPESFM [31]. Table 6 presents reprojection errors, rotation errors and translation errors for models trained without data augmentation, and Table 7 presents corresponding results with data augmentation. The *Inference* and *Inference BA* results are the same as presented in Section 4.2, but now presented side-by-side with the results of 1000 iterations of fine-tuning (the *Fine-tune* column), followed by bundle adjustment (column *Fine-tune + BA*). During fine-tuning, the same loss function is used as during training, but now minimized on the test scene. For both our method and DPESFM [31], fine-tuning can be used to improve the precision of the reconstruction further, and the results of our model trained with data augmentation and succeeded by bundle adjustment are then on par with Colmap.

D.2. Projective Reconstruction of Novel Scenes

In addition to the Euclidean reconstruction results presented in Section 4.2, in Table 8 we also present corresponding results for projective reconstruction, along with a comparison to DPESFM [31] and VarPro [16]. In Table 9, the respective execution times are reported. In the projective setting we also feed normalized image point correspondences as input to the model but, like [31], use Hartley normalization [13] instead of the intrinsic camera parameters. Note that no data augmentation is applied in the projective setting. While doing so with random homography transformations would be feasible, determining a distribution over transformations would not be geometrically interpretable in the same way as for the Euclidean setting, and thus less straightforward.
It can be noted that both our method and DPESFM perform worse at inference of test scenes in the projective setting, as compared to the Euclidean one, but for both methods bundle

		Inference		Inference + BA		Fine-tune		Fine-tune + BA		
		Ours	DPESFM	Ours	DPESFM	Ours	DPESFM	Ours	DPESFM	Colmap
Reprojection error (px)	Alcatraz Courtyard	68.41	68.50	1.70	**0.81** (0.82)	2.90	5.92	**0.81**	**0.81** (**0.81**)	**0.81**
	Alcatraz Water Tower	36.30	50.47	1.05	1.13 (**0.55**)	5.83	147.83	**0.55**	8.63 (**0.55**)	**0.55**
	Drinking Fountain Somewhere in Zurich	36.02	45.87	0.55	**0.31** (7.21)	0.62	27.43	**0.31**	6.73 (**0.31**)	**0.31**
	Nijo Castle Gate	65.71	64.53	**0.73**	**0.73** (5.81)	3.82	5.22	**0.73**	**0.73** (**0.73**)	**0.73**
	Porta San Donato Bologna	74.32	94.26	**0.74**	**0.74** (1.10)	4.42	9.37	**0.74**	**0.74** (0.79)	0.75
	Round Church Cambridge	61.18	55.51	**0.39**	**0.39** (0.50)	3.92	5.92	**0.39**	1.50 (1.51)	**0.39**
	Smolny Cathedral St Petersburg	156.16	120.85	**0.81**	**0.81** (15.15)	3.33	6.68	**0.81**	**0.81** (**0.81**)	**0.81**
	Some Cathedral in Barcelona	150.01	146.10	10.42	12.71 (21.46)	7.99	27.80	**0.89**	**0.89** (**0.89**)	**0.89**
	Sri Veeramakaliamman Singapore	121.02	157.39	2.19	16.87 (16.92)	25.06	40.76	1.20	0.88 (17.26)	**0.71**
	Yueh Hai Ching Temple Singapore	37.46	52.59	**0.65**	**0.65** (1.16)	2.84	8.27	**0.65**	**0.65** (**0.65**)	**0.65**
	Average	80.66	85.61	1.92	3.52 (7.07)	6.07	28.52	0.71	2.24 (2.43)	**0.66**
Rotation error (deg)	Alcatraz Courtyard	10.102	13.201	2.607	**0.035**	1.822	1.423	0.038	0.038	0.043
	Alcatraz Water Tower	10.637	11.053	0.499	0.764	4.161	18.217	0.228	22.764	**0.228**
	Drinking Fountain Somewhere in Zurich	15.846	16.014	0.003	**0.001**	0.453	19.678	**0.001**	22.776	0.007
	Nijo Castle Gate	16.751	10.546	**0.062**	**0.062**	2.301	3.326	0.064	0.064	0.064
	Porta San Donato Bologna	23.839	24.120	0.095	**0.094**	3.664	3.410	**0.094**	**0.094**	0.099
	Round Church Cambridge	18.906	14.473	0.029	**0.026**	6.720	3.116	0.028	1.089	0.035
	Smolny Cathedral St Petersburg	19.387	17.971	0.023	**0.022**	2.969	2.311	0.023	0.023	0.029
	Some Cathedral in Barcelona	27.270	30.471	10.009	20.050	3.874	18.440	**0.019**	**0.019**	0.025
	Sri Veeramakaliamman Singapore	28.275	36.903	0.549	4.871	15.083	30.726	0.218	0.184	**0.169**
	Yueh Hai Ching Temple Singapore	15.733	22.706	**0.038**	**0.038**	3.392	4.624	**0.038**	**0.038**	0.043
	Average	18.675	19.746	1.391	2.596	4.444	10.527	0.075	4.709	**0.074**
Translation error (m)	Alcatraz Courtyard	4.82	4.93	1.09	**0.01**	0.49	0.48	**0.01**	**0.01**	**0.01**
	Alcatraz Water Tower	8.66	7.35	0.31	0.44	2.19	8.61	**0.12**	9.60	**0.12**
	Drinking Fountain Somewhere in Zurich	4.44	4.44	**0.00**	**0.00**	0.04	2.58	**0.00**	1.38	**0.00**
	Nijo Castle Gate	5.07	3.08	**0.01**	**0.01**	0.49	0.79	**0.01**	**0.01**	**0.01**
	Porta San Donato Bologna	9.50	10.72	**0.05**	**0.05**	0.70	0.83	**0.05**	**0.05**	**0.05**
	Round Church Cambridge	8.94	7.19	**0.01**	**0.01**	1.86	1.41	**0.01**	0.56	**0.01**
	Smolny Cathedral St Petersburg	2.70	2.43	**0.01**	**0.01**	0.14	0.24	**0.01**	**0.01**	**0.01**
	Some Cathedral in Barcelona	12.64	12.69	3.22	6.38	1.30	7.72	**0.01**	**0.01**	**0.01**
	Sri Veeramakaliamman Singapore	4.94	4.90	0.16	1.32	3.07	4.65	0.05	0.05	**0.04**
	Yueh Hai Ching Temple Singapore	4.12	4.27	**0.01**	**0.01**	0.41	1.04	**0.01**	**0.01**	**0.01**
	Average	6.58	6.20	0.49	0.82	1.07	2.84	**0.03**	1.17	**0.03**

Table 6. Euclidean reconstruction of novel test scenes, with model trained without data augmentation. The results of DPESFM [31] have been acquired by us training the model, along with the results reported by [31] in parentheses, if available. The result of Colmap, as reported by [31], is also added for reference.

adjustment still converges to relatively good solutions.

D.3. Single-Scene Recovery

While not of major interest to us due to high computational demand, for completeness we also evaluate our model on single-scene recovery, in line with DPESFM [31]. In this setting, the model is "trained" as usual, with the Adam optimizer and the reprojection error loss function with normalized gradients and hinge loss, but on a single scene. This is indeed very similar to bundle adjustment but with a different parameterization. Interestingly, however, Moran et al. [31] found that even with the direct parameterization with free variables for poses and scene points, these modifications to loss function and optimizer alone can result in much better convergence properties than conventional bundle adjustment, when starting from a random initialization. To a large extent, this is probably explained by the presence of the hinge loss to overcome the depth barrier in case of scene points with negative depths.

For these experiments, we have used a slightly shallower model, with $L = 9$ rather than 12 layers, but the feature dimensions and architecture as a whole remain the same. The model is optimized for 100k iterations. Again, warmup is applied by linearly increasing the learning rate from 0 to 1e-4 during the first 2500 iterations, followed by en exponential decay corresponding to a factor of 10 every 35k epochs. Figure 10 shows the resulting average reprojection errors in pixels, compared both with DPESFM [31] as well as other baseline methods reported by [31], i.e. Colmap [36–38], GESFM [24], and Linear [19]. While in many cases our resulting solution has high precision, there are also quite a few failure cases, which are probably cases of suboptimal local reprojection error minima, which interestingly appears to happen more frequently in the single-scene scenario. One possibility is that adding more scenes is effectively flattening the loss landscape, although this should be regarded as nothing more than speculation. In any case, we conclude that our model is more easily trained on multiple scenes simultaneously, in which case we suppose more general geometrical reasoning is encouraged and exploited, and for which our quite expressive model has its edge. Even if we would achieve better performance, it should be noted

		Inference		Inference + BA		Fine-tune		Fine-tune + BA		
		Ours	DPESFM	Ours	DPESFM	Ours	DPESFM	Ours	DPESFM	Colmap
Reprojection error (px)	Alcatraz Courtyard	36.01	92.37	**0.81**	0.92	2.83	4.33	**0.81**	**0.81**	**0.81**
	Alcatraz Water Tower	87.67	2831.94	0.88	10.16	8.65	20.66	**0.55**	0.92	**0.55**
	Drinking Fountain Somewhere in Zurich	219.75	234.90	**0.31**	6.73	0.91	11.40	**0.31**	6.72	**0.31**
	Nijo Castle Gate	61.41	68.19	0.88	0.89	3.58	4.89	**0.73**	**0.73**	**0.73**
	Porta San Donato Bologna	52.15	84.46	0.76	0.75	4.53	8.01	**0.74**	**0.74**	0.75
	Round Church Cambridge	29.80	59.54	**0.39**	1.49	3.23	5.19	**0.39**	1.54	**0.39**
	Smolny Cathedral St Petersburg	85.38	87.81	**0.81**	**0.81**	2.51	3.98	**0.81**	**0.81**	**0.81**
	Some Cathedral in Barcelona	125.68	687.83	**0.89**	16.77	15.73	29.11	**0.89**	1.91	**0.89**
	Sri Veeramakaliamman Singapore	83.50	166.68	2.13	9.30	23.41	43.33	0.80	4.29	**0.71**
	Yueh Hai Ching Temple Singapore	25.60	51.35	**0.65**	0.73	3.15	8.56	**0.65**	**0.65**	**0.65**
	Average	80.69	436.51	0.85	4.86	6.85	13.95	0.67	1.91	**0.66**
Rotation error (deg)	Alcatraz Courtyard	6.093	10.946	0.038	**0.030**	2.293	1.334	0.038	0.035	0.043
	Alcatraz Water Tower	11.501	10.641	0.699	19.351	3.895	5.977	**0.227**	0.668	0.228
	Drinking Fountain Somewhere in Zurich	15.415	15.704	**0.001**	22.759	0.488	21.769	**0.001**	22.747	0.007
	Nijo Castle Gate	17.347	20.032	0.038	**0.036**	1.036	4.537	0.064	0.063	0.064
	Porta San Donato Bologna	18.411	25.004	**0.094**	**0.094**	2.204	6.078	0.097	**0.094**	0.099
	Round Church Cambridge	10.295	18.685	**0.029**	1.086	4.827	3.794	0.030	1.158	0.035
	Smolny Cathedral St Petersburg	11.662	14.380	0.023	**0.019**	2.010	1.170	0.023	0.022	0.029
	Some Cathedral in Barcelona	27.908	29.119	0.020	47.892	16.149	24.625	**0.019**	1.762	0.025
	Sri Veeramakaliamman Singapore	23.702	36.176	0.457	2.759	9.378	33.969	**0.165**	1.035	0.169
	Yueh Hai Ching Temple Singapore	9.515	21.561	**0.038**	**0.038**	4.103	5.770	**0.038**	**0.038**	0.043
	Average	15.185	20.225	0.144	9.406	4.638	10.902	**0.070**	2.762	0.074
Translation error (m)	Alcatraz Courtyard	2.73	5.74	**0.01**	**0.01**	0.60	0.42	**0.01**	**0.01**	**0.01**
	Alcatraz Water Tower	7.53	7.77	0.41	9.05	2.19	3.33	**0.12**	0.40	**0.12**
	Drinking Fountain Somewhere in Zurich	4.45	4.46	**0.00**	1.38	0.08	1.78	**0.00**	1.38	**0.00**
	Nijo Castle Gate	5.67	6.95	**0.01**	**0.01**	0.30	1.01	**0.01**	**0.01**	**0.01**
	Porta San Donato Bologna	4.80	10.48	**0.05**	**0.05**	0.63	1.16	**0.05**	**0.05**	**0.05**
	Round Church Cambridge	5.28	9.00	**0.01**	0.56	1.47	1.48	**0.01**	0.59	**0.01**
	Smolny Cathedral St Petersburg	2.33	2.52	**0.01**	**0.01**	0.11	0.17	**0.01**	**0.01**	**0.01**
	Some Cathedral in Barcelona	12.32	12.66	**0.01**	11.93	6.67	9.82	**0.01**	0.62	**0.01**
	Sri Veeramakaliamman Singapore	4.93	4.90	0.14	0.77	2.28	4.83	**0.04**	0.29	**0.04**
	Yueh Hai Ching Temple Singapore	2.44	4.28	**0.01**	**0.01**	0.49	1.32	**0.01**	**0.01**	**0.01**
	Average	5.25	6.88	0.07	2.38	1.48	2.53	**0.03**	0.34	**0.03**

Table 7. Euclidean reconstruction of novel test scenes, with model trained with data augmentation. The results of DPESFM [31] have been acquired by us training the model. The result of Colmap, as reported by [31], is also added for reference.

	Inference		Inference + BA		Fine-tune		Fine-tune + BA		
	Ours	DPESFM	Ours	DPESFM	Ours	DPESFM	Ours	DPESFM	VarPro
Alcatraz Water Tower	100.16	80.94	1.71	3.86 (7.37)	3.48	9.83	**0.47**	0.99 (**0.47**)	**0.47**
Dinosaur 319	72.80	15.73	1.63	1.22 (1.58)	38.98	3.84	5.88	1.32 (1.30)	**0.43**
Dinosaur 4983	61.54	24.89	1.23	1.10 (3.99)	4.83	4.49	0.64	0.92 (1.14)	**0.42**
Drinking Fountain Somewhere in Zurich	136.73	238.87	**0.28**	**0.28** (14.39)	0.64	2.18	**0.28**	**0.28** (**0.28**)	**0.28**
Eglise Du Dome	60.29	51.72	2.49	1.15 (2.10)	4.66	7.26	0.45	**0.29** (1.27)	-
Gustav Vasa	195.08	525.96	1.63	1.62 (6.30)	1.59	3.12	**0.16**	**0.16** (**0.16**)	**0.16**
Nijo Castle Gate	98.42	80.07	0.40	0.41 (3.27)	125.86	4.11	2.57	**0.39** (**0.39**)	**0.39**
Skansen Kronan Gothenburg	51.23	38.77	0.48	0.70 (1.64)	3.42	2.74	**0.41**	**0.41** (**0.41**)	-
Some Cathedral in Barcelona	452.21	139.54	1.68	1.08 (14.87)	15.35	17.23	1.51	1.67 (**0.51**)	-
Sri Veeramakaliamman Singapore	247.45	165.56	4.40	5.42 (18.25)	190.05	41.00	4.02	**3.85** (5.45)	-
Average	147.59	136.21	1.59	1.68 (7.38)	38.89	9.58	1.64	**1.03** (1.14)	-

Table 8. Reprojection errors of projective reconstruction of novel test scenes, with model trained without data augmentation. The results of DPESFM [31] have been acquired by us training the model, along with the results reported by [31] in parentheses, if available. The result of VarPro, as reported by [31], is also added for reference.

that Colmap works very well, and is much faster to execute than training a model from scratch, albeit not as fast as our learned model combined with bundle adjustment. In Tables 11 and 12, the corresponding rotation and translation errors are reported. Finally, Table 13 shows corresponding results for the projective setting, in which case the results of all baseline methods are again taken as reported by [31].

D.4. Visualizations of Novel Scene Reconstructions

In Figures 6-15, we provide visualizations for the Euclidean reconstructions of all 10 novel test scenes, corresponding to the results reported in Sections 4.2 and4.3, both for our

Scene	Infer.	BA	GPSFM
Alcatraz Water Tower	0.12	84.67	137.06
Dinosaur 319	0.07	1.43	3.25
Dinosaur 4983	0.07	6.11	4.99
Eglise Du Dome	0.42	70.91	105.84
Drinking Fountain Somewhere in Zurich	0.07	3.48	3.35
Gustav Vasa	0.08	4.46	3.45
Nijo Castle Gate	0.13	5.89	6.37
Skansen Kronan Gothenburg	0.27	89.31	93.83
Some Cathedral in Barcelona	0.25	91.21	110.49
Sri Veeramakaliamman Singapore	0.63	133.43	301.71

Table 9. Runtime (s) of our model for projective reconstruction on test scenes, in comparison with VarPro (measured by [31]).

model as well as DPESFM [31]. All point clouds are predictions, together with estimated (red) and ground truth (blue) camera poses. Some outlier filtering has been carried out on the plotted point clouds by coordinate-wise quantiles. Again, please note that the ground truth poses are aligned with the predicted cameras according to Section C, and as such their relative motion is only meaningful if the predicted poses are good enough for the alignment to be accurately estimated (typically after BA but not always before). It can be observed from the plots that our method combined with BA can typically recover high quality scene structures and camera poses, especially when combined with data augmentation. While DPESFM [31] combined with BA also often works relatively well, one notable failure case is *Sri Veeramakaliamman Singapore*, which is a very large scene, that we however manage to recover quite descently. Also note that while our results are in general improved by incorporating data augmentation, it is quite noticeable that DPESFM breaks down from this, possibly due to a limited model capacity / expressivity. In particular, with the help of data augmentation, we are able to recover the *Some Cathedral in Barcelona* scene very well, while DPESFM still fails, and in general shows deterioated results whenever data augmentation is applied.

	Before BA				After BA				
	Ours	DPESFM	GESFM	Linear	Ours	DPESFM	GESFM	Linear	Colmap
Alcatraz Courtyard	**0.98**	1.64	66.5	16.58	**0.81**	**0.81**	4.67	1.27	**0.81**
Alcatraz Water Tower	**1.69**	2.13	131.81	56.26	0.93	**0.55**	25.93	73.72	**0.55**
Buddah Tooth Relic Temple Singapore	**1.73**	2.06	89.94	47.5	**0.85**	**0.85**	13.22	2.66	**0.85**
Doge Palace Venice	**1.28**	3.62	123.53	-	**0.98**	1.00	22.32	-	**0.98**
Door Lund	11.50	**0.32**	(227.0)	20.89	9.94	**0.30**	(9.21)	**0.30**	**0.30**
Drinking Fountain Somewhere in Zurich	0.36	**0.33**	(0.94)	0.58	**0.31**	**0.31**	(0.27)	**0.31**	**0.31**
East Indiaman Goteborg	**1.40**	4.13	170.63	(94.46)	**0.89**	1.85	32.37	(312.9)	**0.89**
Ecole Superior De Guerre	0.54	0.72	(**0.35**)	1.48	0.34	0.34	(**0.14**)	0.34	0.34
Eglise Du Dome	**0.63**	0.91	(90.83)	26.4	**0.27**	**0.27**	(6.21)	0.76	**0.27**
Folke Filbyter	41.44	10.37	(**5.74**)	72.06	11.11	4.29	(0.41)	6.06	**0.29**
Fort Channing Gate Singapore	**0.29**	0.52	2.57	22.69	**0.25**	**0.25**	**0.25**	0.45	**0.25**
Golden Statue Somewhere in Hong Kong	0.49	**0.40**	4.98	73.7	**0.27**	**0.27**	**0.27**	0.3	**0.27**
Gustav II Adolf	15.31	13.91	**6.49**	31.08	11.71	11.49	**0.26**	**0.26**	**0.26**
Gustav Vasa	3.72	**3.52**	(5.21)	11.99	3.15	3.15	(**0.31**)	0.48	0.48
Jonas Ahlstromer	**10.09**	10.82	(36.48)	236.41	7.25	8.41	(0.69)	4.69	**0.22**
King´s College University of Toronto	**0.55**	0.90	(11.87)	(27.29)	**0.34**	**0.34**	(0.35)	(7.12)	**0.34**
Lund University Sphinx	**0.75**	4.78	7.19	60.64	**0.39**	1.36	0.4	4.58	**0.39**
Nijo Castle Gate	**1.69**	1.70	11.18	154.96	**0.73**	**0.73**	**0.73**	4.84	**0.73**
Pantheon Paris	**0.65**	1.47	79.24	39.69	**0.49**	**0.49**	9.71	0.82	-
Park Gate Clermont Ferrand	7.71	**0.57**	1.71	10.5	6.75	**0.35**	**0.35**	**0.35**	**0.35**
Plaza De Armas Santiago	**7.36**	7.40	146.56	-	4.75	4.90	15.61	-	1.13
Porta San Donato Bologna	**1.12**	2.28	29.5	46.12	**0.74**	0.75	3.23	1.16	0.75
Round Church Cambridge	**2.52**	2.66	19.04	9.6	1.50	1.54	2.03	0.41	**0.39**
Skansen Kronan Gothenburg	**0.74**	1.24	8.82	(18.49)	**0.67**	**0.67**	**0.67**	(0.69)	**0.67**
Smolny Cathedral St Petersburg	**0.93**	1.66	19.01	-	**0.81**	**0.81**	1.0	-	**0.81**
Some Cathedral in Barcelona	**1.04**	2.87	47.12	66.97	**0.89**	**0.89**	1.09	2.09	**0.89**
Sri Mariamman Singapore	**1.36**	4.13	52.13	37.16	**0.89**	0.91	7.4	1.17	**0.89**
Sri Thendayuthapani Singapore	**0.87**	23.37	(15.93)	19.57	0.67	8.44	(**0.56**)	0.72	0.67
Sri Veeramakaliamman Singapore	**2.00**	3.47	(205.96)	18.08	**0.71**	0.73	(34.72)	2.2	**0.71**
Statue Of Liberty	113.76	**26.16**	(1031.8)	133.81	32.51	6.97	(52.05)	5.08	**1.25**
The Pumpkin	13.07	33.41	**9.71**	(122.54)	8.67	24.85	**0.57**	(24.19)	**0.57**
Thian Hook Keng Temple Singapore	5.18	**2.75**	53.79	62.7	1.87	1.13	3.32	4.92	**1.12**
Tsar Nikolai I	15.40	9.79	**5.19**	32.86	10.48	6.53	**0.33**	**0.33**	**0.33**
Urban II	17.58	**9.38**	31.71	176.19	12.70	6.92	0.72	17.61	**0.38**
Vercingetorix	6.96	**5.08**	15.87	65.57	5.02	1.50	0.54	2.93	**0.23**
Yueh Hai Ching Temple Singapore	**0.87**	0.94	(27.32)	45.19	**0.65**	**0.65**	(1.64)	2.06	**0.65**

Table 10. Results of single-scene recovery of Euclidean scenes (average reprojection errors in pixels), compared with DPESFM [31], GESFM [24], Linear [19], and Colmap [36–38]. Parentheses mark scenes for which a baseline method has disregarded at least 10% of the cameras.

	Before BA				After BA				
	Ours	DPESFM	GESFM	Linear	Ours	DPESFM	GESFM	Linear	Colmap
Alcatraz Courtyard	**0.424**	0.619	1.851	0.729	0.038	0.049	0.533	**0.042**	0.043
Alcatraz Water Tower	1.763	**0.933**	1.136	1.525	0.677	0.230	9.997	1.525	**0.228**
Buddah Tooth Relic Temple Singapore	1.545	**1.030**	2.95	2.058	**0.081**	**0.081**	4.709	0.551	0.083
Doge Palace Venice	**0.345**	1.163	2.75	-	0.048	0.211	5.317	-	**0.031**
Door Lund	13.940	**0.024**	(2.041)	1.148	12.844	0.006	(7.552)	**0.005**	**0.005**
Drinking Fountain Somewhere in Zurich	0.119	**0.031**	(0.054)	0.077	**0.001**	0.007	(0.01)	0.007	0.007
East Indiaman Goteborg	**1.426**	3.814	11.129	(3.284)	**0.251**	3.117	12.396	(3.284)	**0.251**
Ecole Superior De Guerre	0.243	0.318	(**0.057**)	0.182	**0.018**	0.024	(0.035)	0.024	0.024
Eglise Du Dome	**0.801**	0.808	(2.851)	0.903	**0.031**	0.037	(3.631)	0.162	0.036
Folke Filbyter	74.307	74.596	(**0.332**)	1.94	68.096	70.157	(0.148)	4.484	**0.036**
Fort Channing Gate Singapore	**0.063**	0.207	0.295	0.659	0.010	**0.020**	**0.02**	0.029	**0.020**
Golden Statue Somewhere in Hong Kong	0.692	**0.292**	0.669	8.264	0.024	0.031	0.03	**0.022**	0.031
Gustav II Adolf	88.777	67.784	**0.435**	1.398	66.266	58.458	**0.021**	**0.021**	**0.021**
Gustav Vasa	39.767	34.181	(**0.841**)	1.658	32.316	32.266	(**0.751**)	0.839	0.841
Jonas Ahlstromer	44.994	50.190	(**1.994**)	10.154	49.640	47.117	(0.082)	5.391	**0.036**
King´s College University of Toronto	1.097	0.989	(**0.645**)	(1.07)	0.083	0.085	(**0.059**)	(4.624)	0.084
Lund University Sphinx	0.806	19.522	**0.738**	3.476	**0.025**	8.752	0.058	5.452	0.033
Nijo Castle Gate	0.959	1.495	**0.399**	2.097	**0.064**	0.069	**0.064**	0.744	**0.064**
Pantheon Paris	0.334	**0.192**	3.766	2.655	**0.038**	0.040	3.208	0.072	-
Park Gate Clermont Ferrand	25.295	0.391	**0.203**	0.296	24.779	**0.049**	**0.049**	**0.049**	**0.049**
Plaza De Armas Santiago	**5.964**	6.782	6.291	-	2.299	2.556	6.344	-	**0.122**
Porta San Donato Bologna	**0.377**	2.153	1.013	1.381	**0.093**	0.095	0.513	0.149	0.099
Round Church Cambridge	2.182	2.451	1.021	**0.634**	1.099	1.107	1.851	**0.033**	0.035
Skansen Kronan Gothenburg	**0.235**	0.736	0.549	(0.679)	**0.017**	0.026	0.025	(0.02)	0.025
Smolny Cathedral St Petersburg	**0.469**	0.554	0.493	-	**0.023**	0.033	0.028	-	0.029
Some Cathedral in Barcelona	**0.180**	0.880	1.519	3.126	**0.019**	0.026	0.031	0.057	0.025
Sri Mariamman Singapore	**1.003**	2.302	1.433	1.615	**0.075**	0.077	2.158	0.083	0.078
Sri Thendayuthapani Singapore	**0.835**	46.269	(1.561)	1.581	**0.137**	44.170	(0.329)	0.138	0.138
Sri Veeramakaliamman Singapore	1.876	2.559	(1.807)	**0.519**	**0.167**	0.175	(3.41)	0.288	0.169
Statue Of Liberty	75.495	46.887	(3.449)	**3.357**	73.142	9.091	(8.281)	2.945	**0.213**
The Pumpkin	12.650	94.672	**2.036**	(4.215)	9.136	98.862	0.092	(3.123)	**0.091**
Thian Hook Keng Temple Singapore	3.691	**0.832**	2.751	3.047	0.386	**0.081**	0.245	0.424	0.084
Tsar Nikolai I	72.322	48.499	**0.475**	1.437	74.349	36.280	**0.018**	**0.018**	**0.018**
Urban II	60.201	47.490	**2.077**	8.951	59.713	48.214	0.175	16.348	**0.107**
Vercingetorix	91.624	69.328	**2.203**	2.365	82.565	17.706	1.431	7.138	**0.048**
Yueh Hai Ching Temple Singapore	0.752	**0.720**	(1.813)	1.92	**0.038**	0.043	(0.075)	0.26	0.043

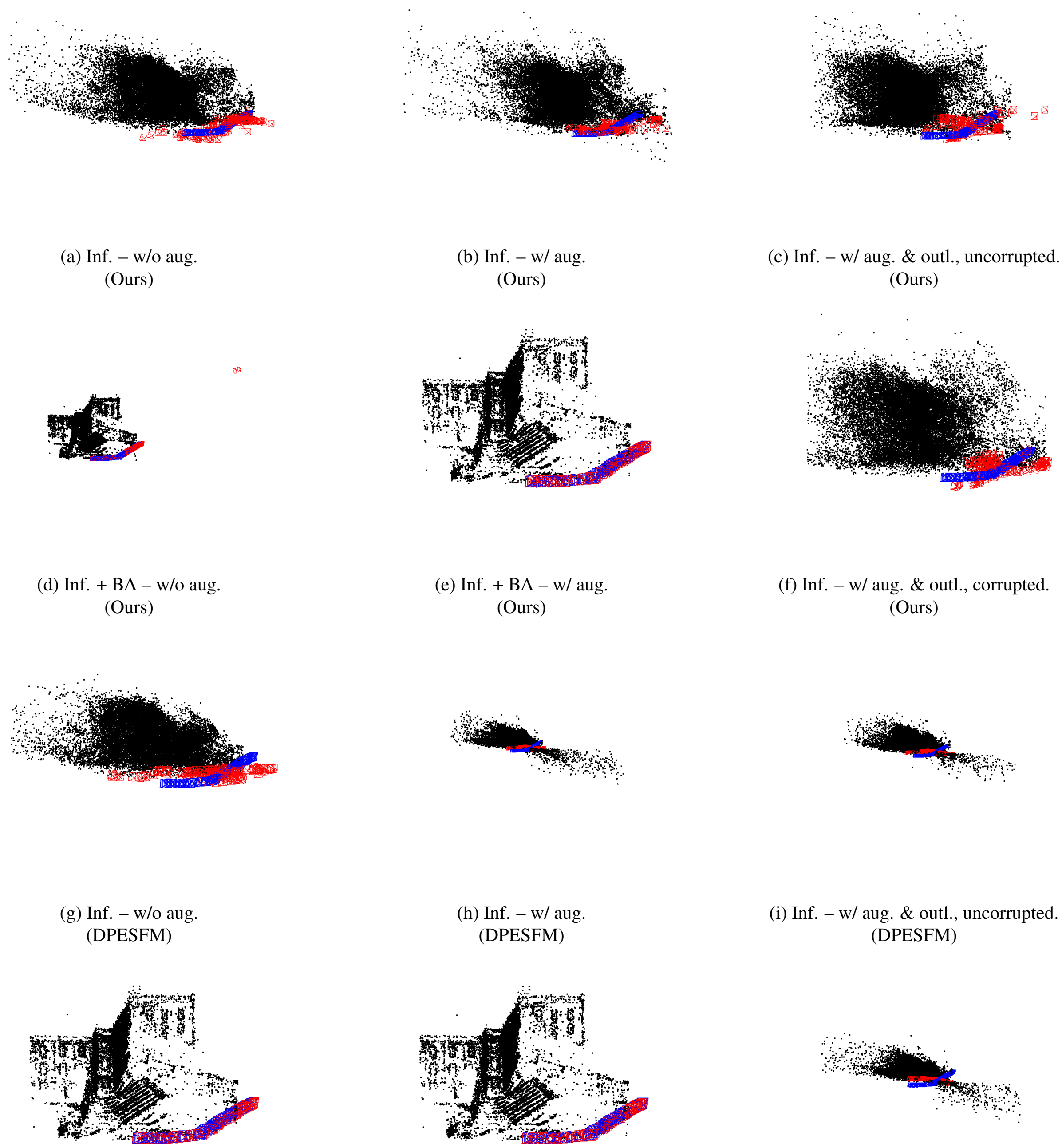
Table 11. Results of single-scene recovery of Euclidean scenes (rotation errors in degrees), compared with DPESFM [31], GESFM [24], Linear [19], and Colmap [36–38]. Parentheses mark scenes for which a baseline method has disregarded at least 10% of the cameras.

	Before BA				After BA				
	Ours	DPESFM	GESFM	Linear	Ours	DPESFM	GESFM	Linear	Colmap
Alcatraz Courtyard	**0.103**	0.160	0.767	0.378	**0.014**	0.015	0.259	**0.014**	**0.014**
Alcatraz Water Tower	1.088	**0.518**	8.332	1.643	0.393	0.116	9.147	1.643	**0.115**
Buddah Tooth Relic Temple Singapore	0.313	**0.233**	2.124	1.325	0.015	**0.014**	1.429	0.125	0.015
Doge Palace Venice	**0.084**	0.342	1.688	-	0.014	0.029	1.608	-	**0.012**
Door Lund	1.009	**0.006**	(1.603)	0.226	1.386	**0.001**	(0.973)	**0.001**	**0.001**
Drinking Fountain Somewhere in Zurich	0.012	**0.004**	(0.016)	0.024	**0.002**	**0.002**	(**0.002**)	**0.002**	**0.002**
East Indiaman Goteborg	**0.348**	0.621	2.783	(2.235)	**0.065**	0.509	3.099	(2.235)	**0.065**
Ecole Superior De Guerre	0.066	0.081	(**0.006**)	0.048	0.005	0.005	(**0.002**)	0.005	0.005
Eglise Du Dome	0.212	0.205	(1.958)	**0.128**	**0.010**	**0.010**	(1.425)	0.046	**0.010**
Folke Filbyter	0.123	0.125	(**0.003**)	0.021	0.110	0.118	(**0.000**)	0.123	**0.000**
Fort Channing Gate Singapore	**0.027**	0.093	0.092	0.139	**0.00008**	**0.008**	**0.008**	0.013	**0.008**
Golden Statue Somewhere in Hong Kong	0.120	**0.073**	0.118	1.153	**0.004**	**0.004**	**0.004**	**0.004**	**0.004**
Gustav II Adolf	13.324	9.714	**0.134**	0.333	9.717	8.524	**0.004**	**0.004**	**0.004**
Gustav Vasa	1.235	1.085	(**0.079**)	0.266	1.136	1.145	(0.101)	**0.099**	0.1
Jonas Ahlstromer	8.649	10.888	(**0.35**)	0.895	10.405	10.451	(0.01)	1.259	0.011
King´s College University of Toronto	0.159	0.235	(**0.152**)	(1.781)	**0.017**	**0.017**	(**0.005**)	(1.877)	**0.017**
Lund University Sphinx	**0.218**	4.585	0.228	1.199	**0.009**	2.191	0.016	1.512	**0.009**
Nijo Castle Gate	0.195	0.286	**0.141**	0.348	**0.011**	0.012	**0.011**	0.19	**0.011**
Pantheon Paris	**0.029**	0.050	0.867	1.275	**0.005**	**0.005**	0.595	0.011	-
Park Gate Clermont Ferrand	11.706	0.125	**0.083**	0.1	11.391	**0.022**	**0.022**	**0.022**	**0.022**
Plaza De Armas Santiago	2.634	2.944	**2.45**	-	1.252	1.383	2.244	-	**0.048**
Porta San Donato Bologna	**0.097**	0.388	0.949	1.588	**0.046**	**0.046**	0.169	0.067	0.047
Round Church Cambridge	0.926	1.003	0.486	**0.217**	0.570	0.582	0.493	**0.012**	**0.012**
Skansen Kronan Gothenburg	**0.071**	0.226	0.223	(0.234)	0.008	0.008	0.008	(**0.007**)	0.008
Smolny Cathedral St Petersburg	**0.021**	0.051	0.209	-	**0.006**	**0.006**	0.007	-	**0.006**
Some Cathedral in Barcelona	**0.063**	0.315	1.776	1.261	0.011	0.011	**0.013**	0.024	0.010
Sri Mariamman Singapore	**0.244**	0.683	1.758	0.721	**0.023**	**0.023**	0.614	0.025	**0.023**
Sri Thendayuthapani Singapore	**0.154**	3.812	(0.285)	0.375	**0.034**	2.870	(0.053)	**0.034**	**0.034**
Sri Veeramakaliamman Singapore	0.432	0.597	(1.966)	**0.273**	**0.038**	0.040	(1.388)	0.095	**0.038**
Statue Of Liberty	27.130	20.012	(4.55)	**3.031**	28.350	4.122	(4.782)	28.049	**0.099**
The Pumpkin	2.881	14.890	**0.513**	(1.656)	2.223	14.952	**0.022**	(14.862)	**0.022**
Thian Hook Keng Temple Singapore	0.450	**0.082**	0.519	0.404	0.055	**0.008**	0.024	0.043	**0.008**
Tsar Nikolai I	12.843	9.467	**0.219**	0.261	13.990	7.836	**0.005**	**0.005**	**0.005**
Urban II	11.350	9.467	**0.774**	2.044	10.882	9.586	0.036	3.038	**0.021**
Vercingetorix	11.202	8.788	**1.158**	2.786	9.696	3.104	0.3	1.564	**0.011**
Yueh Hai Ching Temple Singapore	0.099	**0.098**	(0.642)	0.303	**0.014**	**0.014**	(0.023)	0.059	**0.014**

Table 12. Results of single-scene recovery of Euclidean scenes (translation errors in meters), compared with DPESFM [31], GESFM [24], Linear [19], and Colmap [36–38]. Parentheses mark scenes for which a baseline method has disregarded at least 10% of the cameras.

	Before BA			After BA				
	Ours	DPESFM	GPSFM	Ours	DPESFM	GPSFM	PPSFM	VarPro
Alcatraz Courtyard	**0.77**	1.55	20.34	**0.51**	0.52	0.52	0.57	0.52
Alcatraz Water Tower	**1.00**	2.18	16.50	**0.47**	**0.47**	0.63	0.59	**0.47**
Alcatraz West Side Gardens	**1.27**	9.54	1007.50	**0.72**	0.76	326.99	1.77	-
Basilica Di San Petronio	**6.78**	7.90	1871.41	1.14	0.96	60.69	**0.63**	-
Buddah Statue	418.74	**18.88**	919.26	7.73	2.93	96.96	**0.41**	-
Buddah Tooth Relic Temple Singapore	**1.77**	4.59	18.53	**0.59**	0.60	0.62	0.71	0.60
Corridor	0.32	**0.30**	0.64	**0.26**	**0.26**	**0.26**	0.27	**0.26**
Dinosaur 319	**1.25**	2.35	4.66	0.93	1.53	**0.43**	0.47	**0.43**
Dinosaur 4983	4.99	1.96	**1.54**	0.95	0.57	**0.42**	0.47	**0.42**
Doge Palace Venice	**1.45**	3.60	170.93	**0.60**	**0.60**	3.52	0.67	-
Drinking Fountain Somewhere in Zurich	**0.32**	0.33	1.29	**0.28**	**0.28**	**0.28**	0.31	**0.28**
East Indiaman Goteborg	27.44	**3.31**	99.38	3.23	0.99	5.11	**0.67**	-
Ecole Superior De Guerre	**0.56**	0.75	1.88	**0.26**	**0.26**	**0.26**	0.28	**0.26**
Eglise Du Dome	**0.95**	1.10	8.41	**0.23**	0.24	0.24	0.25	-
Folke Filbyter	115.59	8.87	**1.78**	8.76	8.58	0.82	**0.33**	277.89
Golden Statue Somewhere in Hong Kong	0.56	**0.35**	0.81	**0.22**	**0.22**	**0.22**	0.24	**0.22**
Gustav II Adolf	196.68	14.77	**5.91**	8.10	5.83	**0.23**	0.24	**0.23**
Gustav Vasa	0.56	**0.23**	1.82	**0.16**	**0.16**	**0.16**	0.17	**0.16**
Jonas Ahlstromer	16.16	**14.38**	28.83	5.37	4.72	**0.18**	0.20	**0.18**
King´s College University of Toronto	**2.19**	2.27	22.89	0.51	0.78	2.35	0.26	**0.24**
Lund University Sphinx	81.34	**3.64**	10.00	0.82	**0.34**	0.45	0.37	**0.34**
Model House	0.49	**0.37**	3.66	**0.34**	**0.34**	1.12	0.40	**0.34**
Nijo Castle Gate	1.21	**0.71**	20.08	**0.39**	**0.39**	**0.39**	0.43	**0.39**
Pantheon Paris	**0.96**	1.75	44.85	**0.48**	0.49	2.85	0.62	-
Park Gate Clermont Ferrand	**0.47**	0.61	13.82	**0.31**	**0.31**	0.32	0.49	**0.31**
Plaza De Armas Santiago	**2.42**	5.10	81.01	0.69	**0.64**	3.14	0.71	-
Porta San Donato Bologna	**0.90**	1.58	33.36	**0.40**	**0.40**	0.61	3.75	**0.40**
Skansen Kronan Gothenburg	**0.59**	1.19	8.90	**0.41**	0.41	0.44	0.44	-
Skansen Lejonet Gothenburg	8.22	10.82	69.81	**1.21**	2.05	7.48	1.28	-
Smolny Cathedral St Petersburg	**0.55**	1.66	83.78	**0.46**	**0.46**	**0.46**	0.50	-
Some Cathedral in Barcelona	**0.82**	3.67	14.77	**0.51**	**0.51**	**0.51**	0.54	-
Sri Mariamman Singapore	**2.39**	7.06	39.89	1.02	**0.61**	0.78	0.85	-
Sri Thendayuthapani Singapore	3.30	**2.12**	13.25	2.08	**0.31**	0.56	0.33	-
Sri Veeramakaliamman Singapore	13.80	**6.47**	99.99	3.98	**0.52**	1.78	0.66	-
The Pumpkin	3516.42	14.45	**8.97**	75.77	**0.38**	**0.38**	0.42	-
Thian Hook Keng Temple Singapore	7.99	**7.59**	26.78	**0.54**	**0.54**	0.55	0.66	**0.54**
Tsar Nikolai I	25.75	**6.04**	13.21	4.03	2.43	0.33	0.31	**0.29**
Urban II	487.12	**16.91**	87.25	23.68	6.84	**0.27**	0.31	3.61

Table 13. Results of single-scene recovery of projective scenes (average reprojection errors in pixels), compared with DPESFM [31], GPSFM [23], and PPSFM [28], and VarPro [16].



(a) Inf. – w/o aug. (Ours)

(b) Inf. – w/ aug. (Ours)

(c) Inf. – w/ aug. & outl., uncorrupted. (Ours)

(d) Inf. + BA – w/o aug. (Ours)

(e) Inf. + BA – w/ aug. (Ours)

(f) Inf. – w/ aug. & outl., corrupted. (Ours)

(g) Inf. – w/o aug. (DPESFM)

(h) Inf. – w/ aug. (DPESFM)

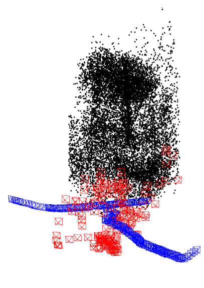
(i) Inf. – w/ aug. & outl., uncorrupted. (DPESFM)

(j) Inf. + BA – w/o aug. (DPESFM)

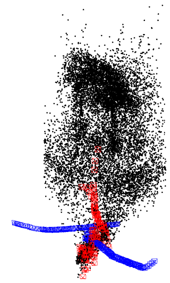
(k) Inf. + BA – w/ aug. (DPESFM)

(l) Inf. – w/ aug. & outl., corrupted. (DPESFM)

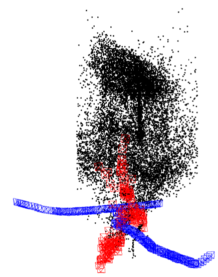
Figure 6. Euclidean reconstruction of *Alcatraz Courtyard*. 6a-6b and 6a-6b show the results of our method with and without data augmentation and bundle adjustment, while 6c and 6f show the results of training with artificially injected outliers, with / without corruption of the test scene as well. Finally, the corresponding results of DPESFM [31] can be seen in 6g-6l.



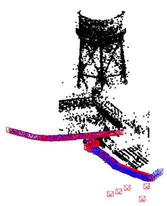
(a) Inf. – w/o aug.
(Ours)



(b) Inf. – w/ aug.
(Ours)

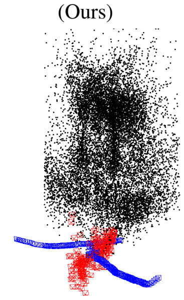


(c) Inf. – w/ aug. & outl., uncorrupted.
(Ours)

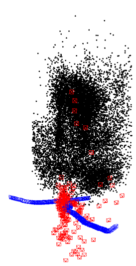


(d) Inf. + BA – w/o aug.
(Ours)

(e) Inf. + BA – w/ aug.
(Ours)

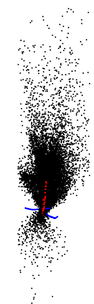


(f) Inf. – w/ aug. & outl., corrupted.
(Ours)

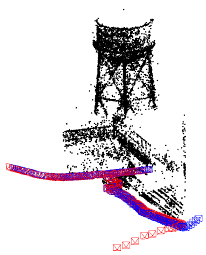


(g) Inf. – w/o aug.
(DPESFM)

(h) Inf. – w/ aug.
(DPESFM)

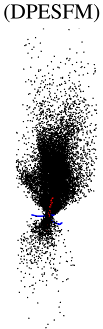


(i) Inf. – w/ aug. & outl., uncorrupted.
(DPESFM)



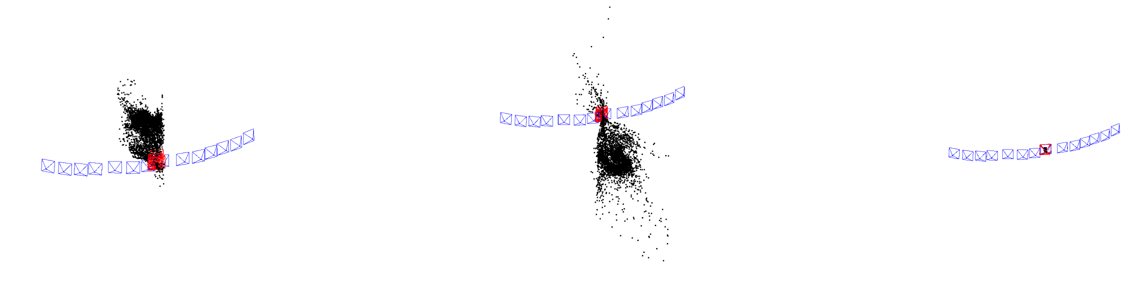
(j) Inf. + BA – w/o aug.
(DPESFM)

(k) Inf. + BA – w/ aug.
(DPESFM)



(l) Inf. – w/ aug. & outl., corrupted.
(DPESFM)

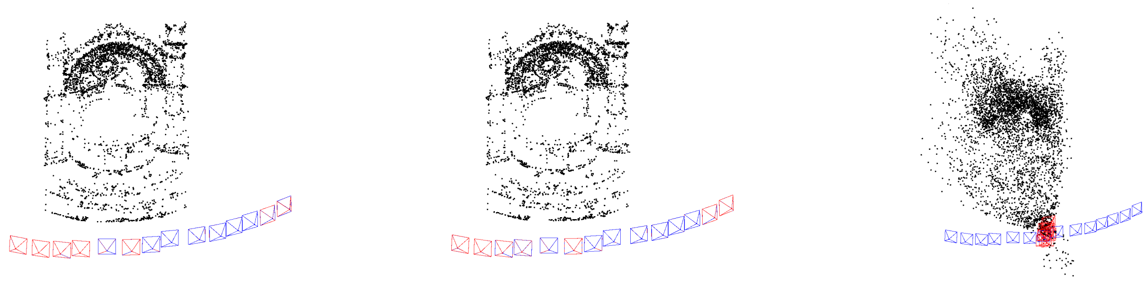
Figure 7. Euclidean reconstruction of *Alcatraz Water Tower*. 7a-7b and 7a-7b show the results of our method with and without data augmentation and bundle adjustment, while 7c and 7f show the results of training with artificially injected outliers, with / without corruption of the test scene as well. Finally, the corresponding results of DPESFM [31] can be seen in 7g-7l.



(a) Inf. – w/o aug.
(Ours)

(b) Inf. – w/ aug.
(Ours)

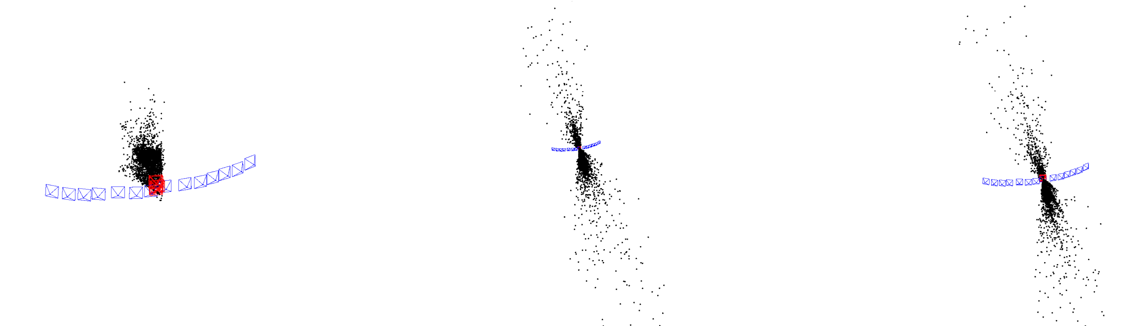
(c) Inf. – w/ aug. & outl., uncorrupted.
(Ours)



(d) Inf. + BA – w/o aug.
(Ours)

(e) Inf. + BA – w/ aug.
(Ours)

(f) Inf. – w/ aug. & outl., corrupted.
(Ours)



(g) Inf. – w/o aug.
(DPESFM)

(h) Inf. – w/ aug.
(DPESFM)

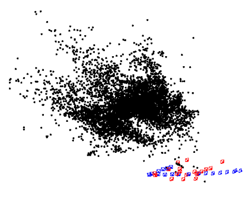
(i) Inf. – w/ aug. & outl., uncorrupted.
(DPESFM)

(j) Inf. + BA – w/o aug.
(DPESFM)

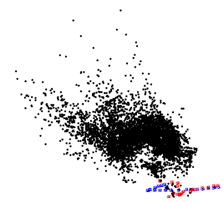
(k) Inf. + BA – w/ aug.
(DPESFM)

(l) Inf. – w/ aug. & outl., corrupted.
(DPESFM)

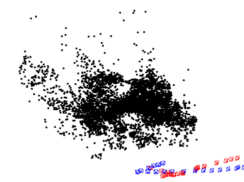
Figure 8. Euclidean reconstruction of *Drinking Fountain Somewhere in Zurich*. 8a-8b and 8a-8b show the results of our method with and without data augmentation and bundle adjustment, while 8c and 8f show the results of training with artificially injected outliers, with / without corruption of the test scene as well. Finally, the corresponding results of DPESFM [31] can be seen in 8g-8l.



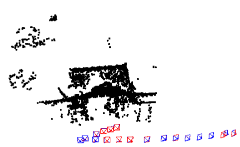
(a) Inf. – w/o aug.
(Ours)



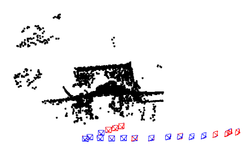
(b) Inf. – w/ aug.
(Ours)



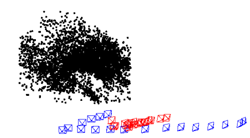
(c) Inf. – w/ aug. & outl., uncorrupted.
(Ours)



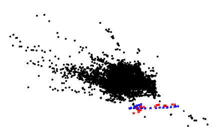
(d) Inf. + BA – w/o aug.
(Ours)



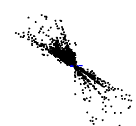
(e) Inf. + BA – w/ aug.
(Ours)



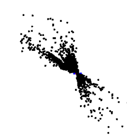
(f) Inf. – w/ aug. & outl., corrupted.
(Ours)



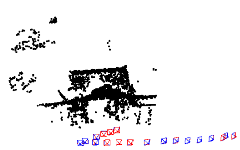
(g) Inf. – w/o aug.
(DPESFM)



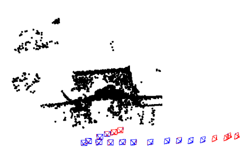
(h) Inf. – w/ aug.
(DPESFM)



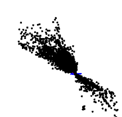
(i) Inf. – w/ aug. & outl., uncorrupted.
(DPESFM)



(j) Inf. + BA – w/o aug.
(DPESFM)

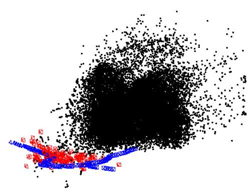


(k) Inf. + BA – w/ aug.
(DPESFM)

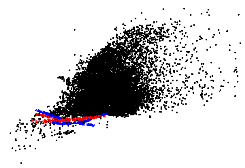


(l) Inf. – w/ aug. & outl., corrupted.
(DPESFM)

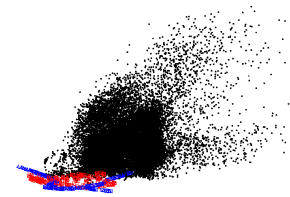
Figure 9. Euclidean reconstruction of *Nijo Castle Gate*. 9a-9b and 9a-9b show the results of our method with and without data augmentation and bundle adjustment, while 9c and 9f show the results of training with artificially injected outliers, with / without corruption of the test scene as well. Finally, the corresponding results of DPESFM [31] can be seen in 9g-9l.



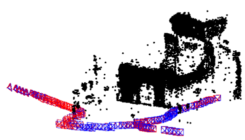
(a) Inf. – w/o aug.
(Ours)



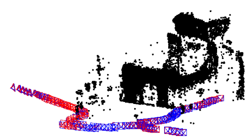
(b) Inf. – w/ aug.
(Ours)



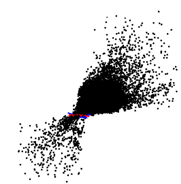
(c) Inf. – w/ aug. & outl., uncorrupted.
(Ours)



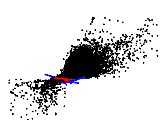
(d) Inf. + BA – w/o aug.
(Ours)



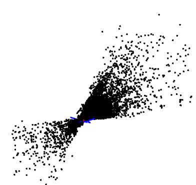
(e) Inf. + BA – w/ aug.
(Ours)



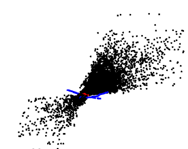
(f) Inf. – w/ aug. & outl., corrupted.
(Ours)



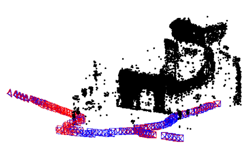
(g) Inf. – w/o aug.
(DPESFM)



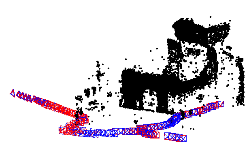
(h) Inf. – w/ aug.
(DPESFM)



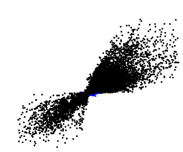
(i) Inf. – w/ aug. & outl., uncorrupted.
(DPESFM)



(j) Inf. + BA – w/o aug.
(DPESFM)

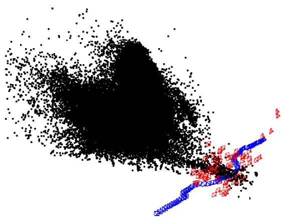


(k) Inf. + BA – w/ aug.
(DPESFM)

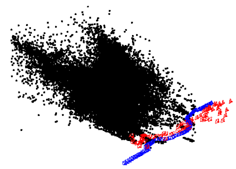


(l) Inf. – w/ aug. & outl., corrupted.
(DPESFM)

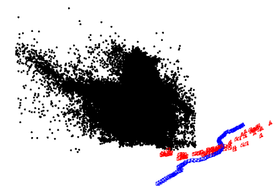
Figure 10. Euclidean reconstruction of *Porta San Donato Bologna*. 10a-10b and 10a-10b show the results of our method with and without data augmentation and bundle adjustment, while 10c and 10f show the results of training with artificially injected outliers, with / without corruption of the test scene as well. Finally, the corresponding results of DPESFM [31] can be seen in 10g-10l.



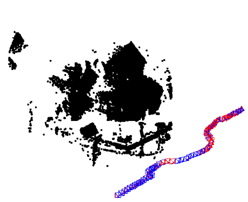
(a) Inf. – w/o aug.
(Ours)



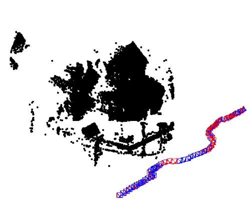
(b) Inf. – w/ aug.
(Ours)



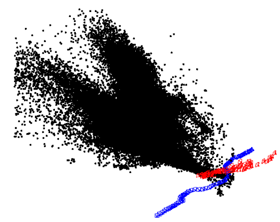
(c) Inf. – w/ aug. & outl., uncorrupted.
(Ours)



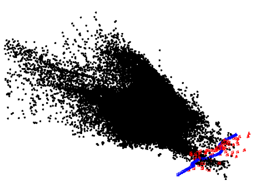
(d) Inf. + BA – w/o aug.
(Ours)



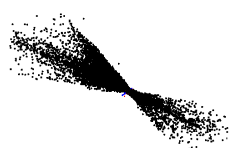
(e) Inf. + BA – w/ aug.
(Ours)



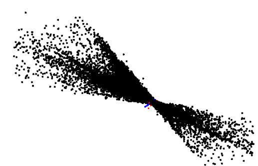
(f) Inf. – w/ aug. & outl., corrupted.
(Ours)



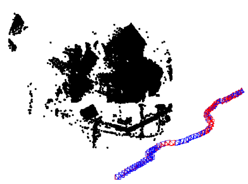
(g) Inf. – w/o aug.
(DPESFM)



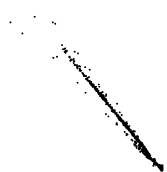
(h) Inf. – w/ aug.
(DPESFM)



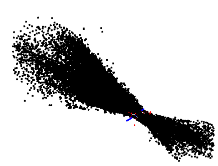
(i) Inf. – w/ aug. & outl., uncorrupted.
(DPESFM)



(j) Inf. + BA – w/o aug.
(DPESFM)

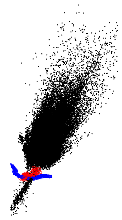


(k) Inf. + BA – w/ aug.
(DPESFM)

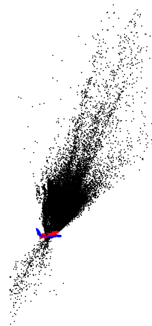


(l) Inf. – w/ aug. & outl., corrupted.
(DPESFM)

Figure 11. Euclidean reconstruction of *Round Church Cambridge*. 11a-11b and 11a-11b show the results of our method with and without data augmentation and bundle adjustment, while 11c and 11f show the results of training with artificially injected outliers, with / without corruption of the test scene as well. Finally, the corresponding results of DPESFM [31] can be seen in 11g-11l.



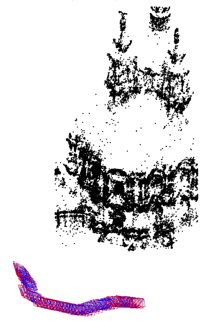
(a) Inf. – w/o aug.
(Ours)



(b) Inf. – w/ aug.
(Ours)

(c) Inf. – w/ aug. & outl., uncorrupted.
(Ours)

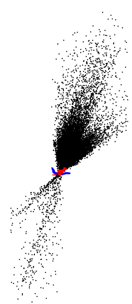
(d) Inf. + BA – w/o aug.
(Ours)



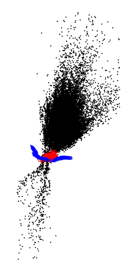
(e) Inf. + BA – w/ aug.
(Ours)

(f) Inf. – w/ aug. & outl., corrupted.
(Ours)

(g) Inf. – w/o aug.
(DPESFM)



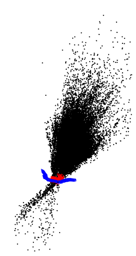
(h) Inf. – w/ aug.
(DPESFM)



(i) Inf. – w/ aug. & outl., uncorrupted.
(DPESFM)

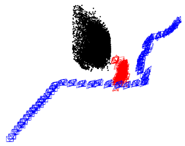
(j) Inf. + BA – w/o aug.
(DPESFM)

(k) Inf. + BA – w/ aug.
(DPESFM)

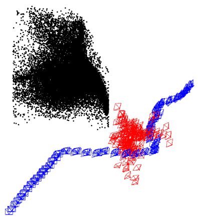


(l) Inf. – w/ aug. & outl., corrupted.
(DPESFM)

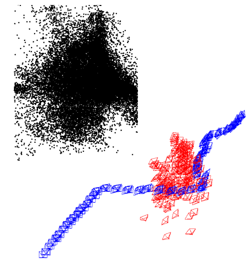
Figure 12. Euclidean reconstruction of *Smolny Cathedral St Petersburg*. 12a-12b and 12a-12b show the results of our method with and without data augmentation and bundle adjustment, while 12c and 12f show the results of training with artificially injected outliers, with / without corruption of the test scene as well. Finally, the corresponding results of DPESFM [31] can be seen in 12g-12l.



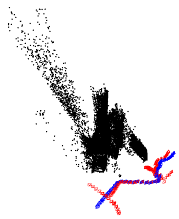
(a) Inf. – w/o aug.
(Ours)



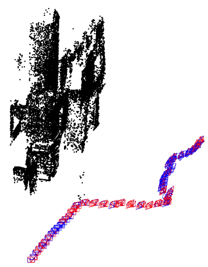
(b) Inf. – w/ aug.
(Ours)



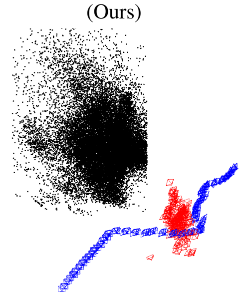
(c) Inf. – w/ aug. & outl., uncorrupted.
(Ours)



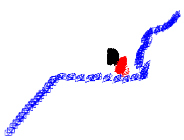
(d) Inf. + BA – w/o aug.
(Ours)



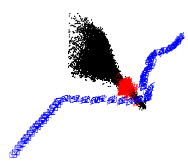
(e) Inf. + BA – w/ aug.
(Ours)



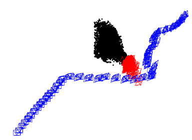
(f) Inf. – w/ aug. & outl., corrupted.
(Ours)



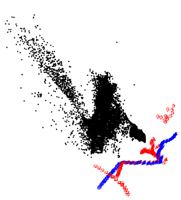
(g) Inf. – w/o aug.
(DPESFM)



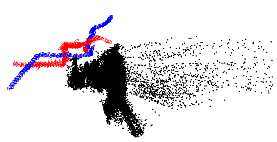
(h) Inf. – w/ aug.
(DPESFM)



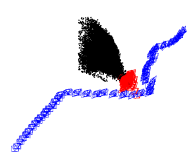
(i) Inf. – w/ aug. & outl., uncorrupted.
(DPESFM)



(j) Inf. + BA – w/o aug.
(DPESFM)

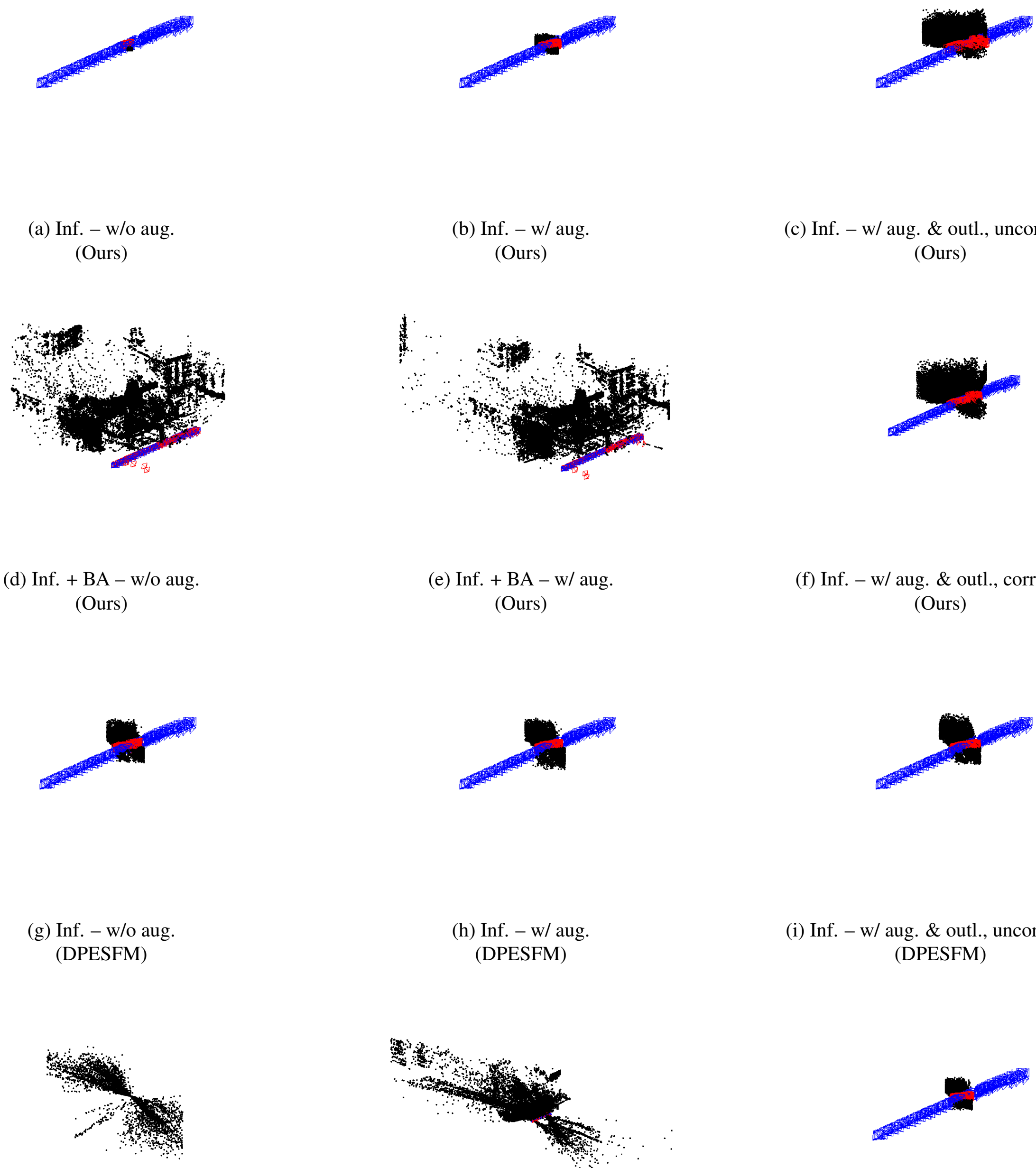


(k) Inf. + BA – w/ aug.
(DPESFM)



(l) Inf. – w/ aug. & outl., corrupted.
(DPESFM)

Figure 13. Euclidean reconstruction of *Some Cathedral in Barcelona*. 13a-13b and 13a-13b show the results of our method with and without data augmentation and bundle adjustment, while 13c and 13f show the results of training with artificially injected outliers, with / without corruption of the test scene as well. Finally, the corresponding results of DPESFM [31] can be seen in 13g-13l.



(a) Inf. – w/o aug.
(Ours)

(b) Inf. – w/ aug.
(Ours)

(c) Inf. – w/ aug. & outl., uncorrupted.
(Ours)

(d) Inf. + BA – w/o aug.
(Ours)

(e) Inf. + BA – w/ aug.
(Ours)

(f) Inf. – w/ aug. & outl., corrupted.
(Ours)

(g) Inf. – w/o aug.
(DPESFM)

(h) Inf. – w/ aug.
(DPESFM)

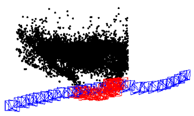
(i) Inf. – w/ aug. & outl., uncorrupted.
(DPESFM)

(j) Inf. + BA – w/o aug.
(DPESFM)

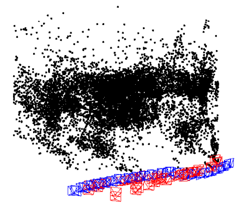
(k) Inf. + BA – w/ aug.
(DPESFM)

(l) Inf. – w/ aug. & outl., corrupted.
(DPESFM)

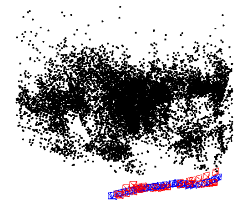
Figure 14. Euclidean reconstruction of *Sri Veeramakaliamman Singapore*. 14a-14b and 14a-14b show the results of our method with and without data augmentation and bundle adjustment, while 14c and 14f show the results of training with artificially injected outliers, with / without corruption of the test scene as well. Finally, the corresponding results of DPESFM [31] can be seen in 14g-14l.



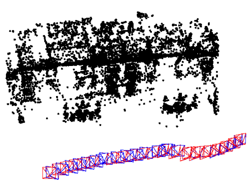
(a) Inf. – w/o aug.
(Ours)



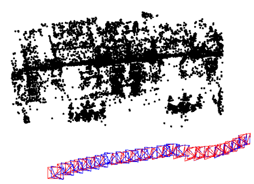
(b) Inf. – w/ aug.
(Ours)



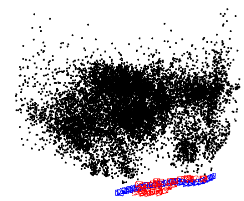
(c) Inf. – w/ aug. & outl., uncorrupted.
(Ours)



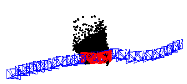
(d) Inf. + BA – w/o aug.
(Ours)



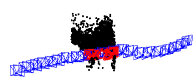
(e) Inf. + BA – w/ aug.
(Ours)



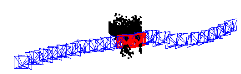
(f) Inf. – w/ aug. & outl., corrupted.
(Ours)



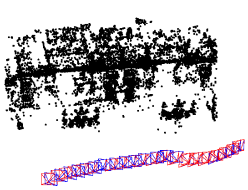
(g) Inf. – w/o aug.
(DPESFM)



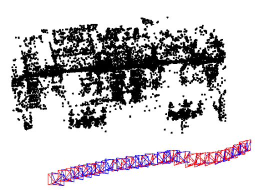
(h) Inf. – w/ aug.
(DPESFM)



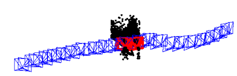
(i) Inf. – w/ aug. & outl., uncorrupted.
(DPESFM)



(j) Inf. + BA – w/o aug.
(DPESFM)



(k) Inf. + BA – w/ aug.
(DPESFM)



(l) Inf. – w/ aug. & outl., corrupted.
(DPESFM)

Figure 15. Euclidean reconstruction of *Yueh Hai Ching Temple Singapore*. 15a-15b and 15a-15b show the results of our method with and without data augmentation and bundle adjustment, while 15c and 15f show the results of training with artificially injected outliers, with / without corruption of the test scene as well. Finally, the corresponding results of DPESFM [31] can be seen in 15g-15l.