Supplementary Material:
NormalLoc: Visual Localization on Textureless 3D Models using Surface Normals.

1. Completed results on the CADLoc benchmark

The CADLoc [1] benchmark provides an evaluation script which measures localization accuracy using four types of metrics. All metrics are based on the Dense Correspondence Reprojection Error (DCRE) [1, 2]. The DCRE for each query is computed as follows: we obtain a 3D point for each pixel in the depth map and project the 3D point into the image coordinates by using the ground truth and the estimated pose, respectively. As a result, we obtain a set of 2D-2D correspondences whose distances represent the difference between the ground truth and the estimated poses. Using the DCRE for each query image, four types of metrics are computed as follows:

- *mean / max. DCRE - GA*: For each query image, the mean or maximum Euclidean distance among the 2D-2D correspondences is computed and normalized so that the length of the image diagonal is 100%. The mean or maximum value is used as the localzation accuracy score for the query image. In the "mean / max. DCRE -GA" metrics, the ground truth poses are computed by aligning CAD models to reference meshes derived from Multi-View-Stereo (MVS) for the query images. These ground truth poses are referred to as Global Alignment (GA) poses.
- *mean / max. DCRE - LR*: For each query image, the mean or maximum Euclidean distance is computed in the same way as for the "mean / max. DCRE -GA" metrics. In the "mean / max. DCRE - LR" metrics, the ground truth poses are computed by aligning depth maps rendered from the reference meshes from the viewpoints of the GA poses to the CAD models. These ground truth poses are referred to as Locally Refined (LR) poses.

Table 1. Performance evaluation on the CADLoc benchmark - on **the Notre Dame (E) model**. Performance is evaluated as the success ratio under three thresholds (10% / 20% / 30%) for each metric.

method	mean DCRE - GA	max. DCRE - GA	mean DCRE - LR	max. DCRE - LR
MeshLoc - LoFTR	70.9 / 75.7 / 82.0	64.6 / 71.4 / 75.1	70.9 / 75.7 / 83.1	64.0 / 71.4 / 74.6
MeshLoc - Patch2Pix+SG	66.1 / 74.6 / 80.4	60.8 / 67.7 / 73.5	65.6 / 75.1 / 80.4	61.4 / 67.2 / 73.5
MeshLoc - SuperGlue	70.9 / 78.3 / 83.1	63.0 / 71.4 / 77.8	70.9 / 78.3 / 82.5	64.0 / 71.4 / 77.2
NormalLoc	**78.8 / 82.5 / 86.2**	**75.7 / 79.9 / 82.0**	**79.4 / 83.1 / 86.8**	**75.7 / 79.9 / 82.0**

Table 2. Performance evaluation on the CADLoc benchmark - on **the Notre Dame (F) model**. Performance is evaluated as the success ratio under three thresholds (10% / 20% / 30%) for each metric.

method	mean DCRE - GA	max. DCRE - GA	mean DCRE - LR	max. DCRE - LR
MeshLoc - LoFTR	5.3 / 12.2 / 21.2	3.2 / 10.1 / 16.4	5.3 / 12.2 / 21.2	3.2 / 10.6 / 16.4
MeshLoc - Patch2Pix+SG	6.9 / 18.0 / 24.3	3.7 / 7.9 / 13.8	6.9 / 17.5 / 23.8	3.2 / 8.5 / 13.2
MeshLoc - SuperGlue	23.8 / 36.5 / 48.1	13.8 / 25.4 / 31.2	23.3 / 36.5 / 48.1	13.8 / 25.4 / 31.2
NormalLoc	**79.9 / 82.5 / 87.3**	**63.5 / 78.8 / 80.4**	**79.9 / 82.5 / 87.3**	**63.5 / 78.8 / 80.4**

Table 3. Performance evaluation on the CADLoc benchmark - on **the Notre Dame (G) model**. Performance is evaluated as the success ratio under three thresholds (10% / 20% / 30%) for each metric.

method	mean DCRE - GA	max. DCRE - GA	mean DCRE - LR	max. DCRE - LR
MeshLoc - LoFTR	1.6 / 10.1 / 20.1	0.5 / 4.2 / 12.7	1.6 / 13.2 / 19.6	1.1 / **6.9** / 13.2
MeshLoc - Patch2Pix+SG	1.6 / 5.8 / 22.2	0.5 / 2.1 / 5.3	1.6 / 5.3 / 20.1	0.5 / 2.6 / 5.8
MeshLoc - SuperGlue	4.8 / 14.3 / 30.7	0.5 / 5.8 / **14.3**	5.8 / 13.8 / 31.7	1.6 / 6.3 / **14.3**
NormalLoc	**5.8 / 21.2 / 39.2**	**2.1 / 6.3** / 10.1	**6.9 / 21.2 / 40.7**	**2.6** / 5.3 / 11.6

Table 4. Performance evaluation on the CADLoc benchmark - on **the Pantheon (C) model**. Performance is evaluated as the success ratio under three thresholds (10% / 20% / 30%) for each metric.

method	mean DCRE - GA	max. DCRE - GA	mean DCRE - LR	max. DCRE - LR
MeshLoc - LoFTR	67.4 / 85.8 / 90.1	51.8 / 65.2 / 75.9	68.1 / 83.7 / 90.1	51.1 / 63.8 / 74.5
MeshLoc - Patch2Pix+SG	17.7 / 27.7 / 46.1	7.1 / 17.0 / 23.4	17.7 / 28.4 / 45.4	7.1 / 17.0 / 24.1
MeshLoc - SuperGlue	29.8 / 46.1 / 57.4	7.1 / 24.8 / 36.2	29.1 / 46.8 / 58.2	5.7 / 24.8 / 36.9
NormalLoc	**95.7 / 99.3 / 100.0**	**63.8 / 81.6 / 85.8**	**95.7 / 97.9 / 100.0**	**59.6 / 82.3 / 85.8**

Table 5. Performance evaluation on the CADLoc benchmark - on **the Pantheon (D) model**. Performance is evaluated as the success ratio under three thresholds (10% / 20% / 30%) for each metric.

method	mean DCRE - GA	max. DCRE - GA	mean DCRE - LR	max. DCRE - LR
MeshLoc - LoFTR	9.2 / 18.4 / 33.3	1.4 / 12.8 / 21.3	9.2 / 19.1 / 34.8	1.4 / 13.5 / 22.7
MeshLoc - Patch2Pix+SG	25.5 / 39.0 / 48.9	17.0 / 27.0 / 34.8	25.5 / 39.7 / 48.9	17.7 / 27.0 / 34.8
MeshLoc - SuperGlue	38.3 / 53.2 / 61.0	20.6 / 38.3 / 48.2	39.0 / 53.2 / 61.0	19.9 / 39.0 / 49.6
NormalLoc	**97.9 / 98.6 / 99.3**	**84.4 / 97.2 / 97.9**	**97.2 / 98.6 / 99.3**	**75.2 / 95.7 / 97.2**

Table 6. Performance evaluation on the CADLoc benchmark - on **the Pantheon (E) model**. Performance is evaluated as the success ratio under three thresholds (10% / 20% / 30%) for each metric.

method	mean DCRE - GA	max. DCRE - GA	mean DCRE - LR	max. DCRE - LR
MeshLoc - LoFTR	9.9 / 33.3 / 52.5	0.0 / 10.6 / 24.8	3.5 / 32.6 / 50.4	0.0 / 9.2 / 23.4
MeshLoc - Patch2Pix+SG	16.3 / 22.7 / 34.8	14.2 / 17.7 / 21.3	14.9 / 21.3 / 33.3	14.2 / 17.0 / 20.6
MeshLoc - SuperGlue	4.3 / 20.6 / 37.6	1.4 / 6.4 / 17.0	5.0 / 17.0 / 36.9	2.1 / 4.3 / 15.6
NormalLoc	**91.5 / 96.5 / 98.6**	**52.5 / 92.9 / 94.3**	**90.1 / 95.7 / 98.6**	**34.8 / 88.7 / 94.3**

Table 7. Performance evaluation on the CADLoc benchmark - on **the Reichstag (E) model**. Performance is evaluated as the success ratio under three thresholds (10% / 20% / 30%) for each metric.

method	mean DCRE - GA	max. DCRE - GA	mean DCRE - LR	max. DCRE - LR
MeshLoc - LoFTR	5.3 / 30.7 / 60.0	0.0 / 6.7 / 22.7	5.3 / 32.0 / 58.7	1.3 / 5.3 / 22.7
MeshLoc - Patch2Pix+SG	36.0 / 56.0 / 68.0	13.3 / 33.3 / 42.7	37.3 / 54.7 / 68.0	13.3 / 33.3 / 41.3
MeshLoc - SuperGlue	48.0 / 72.0 / 90.7	16.0 / 42.7 / 61.3	48.0 / 72.0 / 90.7	16.0 / 42.7 / 60.0
NormalLoc	**74.7 / 89.3 / 93.3**	**48.0 / 73.3 / 78.7**	**74.7 / 88.0 / 93.3**	**48.0 / 72.0 / 78.7**

References

[1] Vojtech Panek, Zuzana Kukelova, and Torsten Sattler. Visual localization using imperfect 3d models from the internet. In *CVPR*, pages 13175–13186, 2023. 1

[2] Anirudh Viswanathan, Bernardo R Pires, and Daniel Huber. Vision based robot localization by ground to satellite matching in gps-denied situations. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 192–198. IEEE, 2014. 1

Table 8. Performance evaluation on the CADLoc benchmark - on **the Reichstag (F) model**. Performance is evaluated as the success ratio under three thresholds (10% / 20% / 30%) for each metric.

method	mean DCRE - GA	max. DCRE - GA	mean DCRE - LR	max. DCRE - LR
MeshLoc - LoFTR	5.3 / 18.7 / 30.7	4.0 / 6.7 / 16.0	6.7 / 17.3 / 29.3	2.7 / 5.3 / 14.7
MeshLoc - Patch2Pix+SG	5.3 / 14.7 / 26.7	2.7 / 5.3 / 9.3	5.3 / 14.7 / 25.3	2.7 / 5.3 / 8.0
MeshLoc - SuperGlue	6.7 / 17.3 / 40.0	1.3 / 10.7 / 17.3	6.7 / 17.3 / 40.0	1.3 / 10.7 / 16.0
NormalLoc	**86.7 / 96.0 / 98.7**	**66.7 / 81.3 / 89.3**	**86.7 / 93.3 / 100.0**	**69.3 / 84.0 / 90.7**

Table 9. Performance evaluation on the CADLoc benchmark - on **the St. Peters Square (D) model**. Performance is evaluated as the success ratio under three thresholds (10% / 20% / 30%) for each metric.

method	mean DCRE - GA	max. DCRE - GA	mean DCRE - LR	max. DCRE - LR
MeshLoc - LoFTR	7.9 / 29.4 / 43.7	3.2 / 17.5 / 29.4	1.6 / 11.1 / 19.0	0.8 / 2.4 / **9.5**
MeshLoc - Patch2Pix+SG	15.1 / 37.3 / 49.2	4.0 / 7.9 / 19.0	3.2 / 18.3 / 34.9	**1.6** / 1.6 / 6.3
MeshLoc - SuperGlue	28.6 / 47.6 / 63.5	3.2 / 14.3 / 31.7	5.6 / 24.6 / 48.4	0.8 / 4.8 / **9.5**
NormalLoc	**45.2 / 64.3 / 75.4**	**7.1 / 26.2 / 46.8**	**9.5 / 32.5 / 57.9**	**1.6 / 5.6** / 8.7