Supplementary Material for
DKM: Dense Kernelized Feature Matching for Geometry Estimation

Johan Edstedt, Ioannis Athanasiadis, Mårten Wadenbäck, Michael Felsberg

Computer Vision Laboratory
Linköping University

Table 1. Pose estimation results on the Megadepth-8-Scenes benchmark, measured in AUC (higher is better). Top section, sparse methods, bottom section, dense methods.

Method ↓	AUC →	@5°	@10°	@20°
ASpanFormer [1] ECCV'22		57.2	72.1	82.9
PDCNet+ [4] Arxiv'21		51.8	66.6	77.2
DKM		**60.5**	**74.5**	**84.2**

1. Additional State-of-the-Art Comparison

MegaDepth-8-Scenes Pose Estimation:
Since the MegaDepth-1500 benchmark is sampled from only 2 scenes, it is of interest to ascertain that results hold in a wider setting. We therefore sample a total of 1600 pairs from 8 different scenes:

1. Piazza San Marco (0008): Example in Figure 1.
2. Sagrada Familia (0019): Example in Figure 2.
3. Lincoln Memorial Statue (0021): See main paper.
4. British Museum (0024): Example in Figure 3.
5. Tower of London (0025): See main paper.
6. Florence Cathedral (0032): Example in Figure 4.
7. Milan Cathedral (0063): See main paper.
8. Mount Rushmore (1589): Example in Figure 5.

We use the same protocol as in MegaDepth-1500. We call this new benchmark *MegaDepth-8-Scenes*. Results on this benchmark are presented in Table 1. We achieve state-of-the-art results here as well, with a relative performance increase of +3.3 AUC@5° compared to the previous best sparse method, and by +8.7 percentage points compared to the previous best dense method.
St. Paul's Cathedral: COTR and ECO-TR [2, 3] are two recent dense methods based on transformer architectures.

Figure 1. Qualitative example of pair in Piazza San Marco (0008) with DKM warp and certainty.

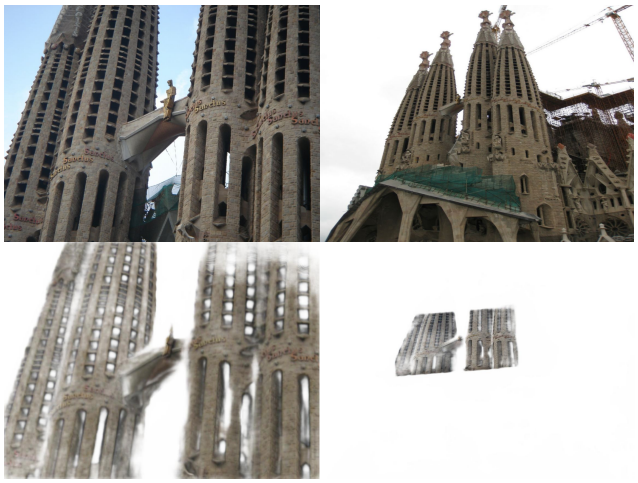

Figure 2. Qualitative example of pair in Sagrada Familia (0019) with DKM warp and certainty.

Figure 3. Qualitative example of pair in Britism Museum (0024) with DKM warp and certainty.

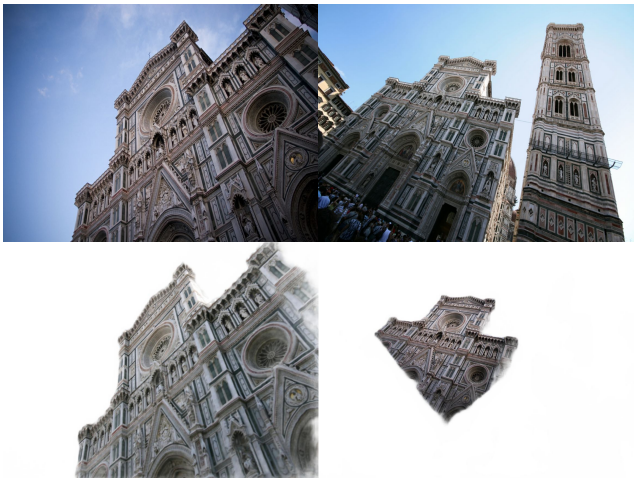

Figure 4. Qualitative example of pair in Florence Cathedral (0032) with DKM warp and certainty.

Here we compare results of our approach compared to those works on the St. Paul's Cathedral scene. We use the evaluation protocol of ECO-TR. We present results in Table 2. We find that our method outperforms both COTR and ECO-TR, achieving a performance increase of +8.0 mAA@5°. We additionally present a representative qualitative example in Figure 6.

2. Further Qualitative Examples

2.1. MegaDepth-1500

In Figure 7 we present a qualitative example on the St. Peter's Basilica (0015) scene.

Figure 5. Qualitative example of pair in Mount Rushmore (1589) with DKM warp and certainty.

Table 2. Pose estimation results on the St. Paul's Cathedral benchmark, measured in mAA (higher is better). We report the average and estimated standard deviation over five runs.

Method ↓	mAA →	@5°	@10°
COTR [2] ICCV'21		44.3	66.0
ECO-TR [3] ECCV'22		45.3	66.1
DKM		**53.3**	**72.1**

Figure 6. Qualitative example of DKM warp and certainty on the St. Paul's Cathedral benchmark.

2.2. HPatches

In Figures 8 and 9 we present qualitative results on HPatches. We find that despite not being trained for planar scenes, DKM performs very well here as well.

Figure 7. DKM warp and certainty on a pair from the St. Peter's Basilica (0015) scene.

Figure 8. DKM result on the HPatches planar scene v_bird.

2.3. ScanNet

In Figure 10, we present a qualiative example of the indoor model of DKM on the ScanNet-1500 benchmark.

3. Additional Failure Cases

Extreme Lack of Texture: In Figure 11 we show a failure case where our method completely fails. We believe this failure is due to the complete lack of unique local textures. However, the matching is not ill-defined as unique global patterns exist. Encouragingly however, the model predicts a very low certainty for this pair, indicating a well calibrated uncertainty estimate.

Figure 9. DKM result on the HPatches planar scene v_graffiti.

Figure 10. DKM indoor model results on a kitchen scene in the ScanNet-1500 benchmark.

Figure 11. Failure case of DKM. The warp completely fails, and the estimated certainty is very low.

References

[1] Hongkai Chen, Zixin Luo, Lei Zhou, Yurun Tian, Mingmin Zhen, Tian Fang, David Mckinnon, Yanghai Tsin, and Long Quan. ASpanFormer: Detector-free image matching with adaptive span transformer. In *Proc. European Conference on Computer Vision (ECCV)*, 2022. 1

[2] Wei Jiang, Eduard Trulls, Jan Hosang, Andrea Tagliasacchi, and Kwang Moo Yi. COTR: Correspondence Transformer for Matching Across Images. In *ICCV*, 2021. 1, 2

[3] Dongli Tan, Jiang-Jiang Liu, Xingyu Chen, Chao Chen, Ruixin Zhang, Yunhang Shen, Shouhong Ding, and Rongrong Ji. ECO-TR: Efficient Correspondences Finding Via Coarse-to-Fine Refinement. In *Proc. European Conference on Computer Vision (ECCV)*, 2022. 1, 2

[4] Prune Truong, Martin Danelljan, Radu Timofte, and Luc Van Gool. Pdc-net+: Enhanced probabilistic dense correspondence network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–18, 2023. 1