Supplementary material of "Geometrized Transformer for Self-Supervised Homography Estimation"

Jiazhen Liu and Xirong Li
Key Lab of DEKE, Renmin University of China
https://github.com/ruc-aimc-lab/GeoFormer

In this supplementary material, we report additional experimental results which are not included in the main paper due to space limits.

Table 7 shows the performance of varied methods on HPatches and ISC-HE. Table 8 shows the HPatches results divided between "Viewpoint" and "Illumination" splits. Visual localization results on the InLoc benchmark [15] are given in Table 9.

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 *ECCV*, 2022. 1

Table 7: **Performance of different methods for homography estimation**. Per method, we use the star symbol (*) to indicate its official release trained on the author-preferred dataset, while the no-star counterpart is re-trained by us on Oxford-Paris, using author-provided code.

Method	Testset: HPatches				Testset: ISC-HE			
	@3px	@5px	@10px	mAUC	@3px	@5px	@10px	mAUC
Deep Homography:								
CA-Unsupervised* [8], trained on custom dataset	20.5	31.7	40.1	30.8	8.9	16.4	23.4	16.2
CA-Unsupervised	37.6	41.9	45.3	41.6	10.7	17.5	23.8	17.3
HomoGAN* [5], trained on the same dataset as [8]	34.2	38.3	42.1	38.2	9.7	15.9	22.4	16.0
HomoGAN	31.3	36.8	41.8	36.6	9.1	15.6	22.3	15.7
BasesHomo* [16], trained on the same dataset as [8]	38.3	42.4	45.5	42.1	9.9	17.6	25.0	17.5
BasesHomo	39.6	43.2	45.9	42.9	10.1	17.1	23.9	17.0
Detector-based matching:								
SuperPoint* [2], trained on MS-COCO 2014 [7]	43.4	57.6	72.7	57.9	18.3	39.0	62.2	39.8
SuperPoint	42.1	56.2	69.9	56.1	16.2	37.8	61.4	38.5
R2D2* [9], trained on Aachen Day-Night dataset [13, 12]	50.6	63.9	76.8	63.8	18.2	39.6	62.9	40.2
R2D2	43.4	56.6	71.7	57.2	18.7	40.0	61.9	40.2
SuperGlue* [11], trained on MegaDepth [6]	53.9	68.3	81.7	68.0	19.6	42.2	66.9	42.9
SuperGlue	45.2	60.1	76.1	60.5	19.2	41.7	65.8	42.2
Detector-free matching:								
NCNet* [10], trained on PF-Pascal dataset [4]	48.3	50.1	59.8	52.7	9.6	25.3	51.2	28.7
NCNet	43.2	48.3	56.9	49.5	9.5	25.7	49.3	28.2
LoFTR*[14], trained on MegaDepth	65.3	75.2	84.6	75.0	18.6	38.8	60.5	39.3
LoFTR	58.5	69.8	81.1	69.8	18.7	41.0	64.8	41.5
ASpanFormer* [1], trained on MegaDepth	67.0	76.9	85.6	76.5	19.3	41.1	63.8	41.4
ASpanFormer	59.9	71.1	81.6	70.9	18.0	39.2	62.0	39.7
DKM* [3], trained on MegaDepth	70.7	**80.2**	**88.4**	79.8	19.1	40.4	63.4	41.0
DKM [3]	30.6	37.3	44.5	38.1	7.1	15.3	25.6	16.0
GeoFormer, trained on MegaDepth	**72.1**	80.0	87.7	**79.9**	**20.5**	**44.5**	**69.0**	**44.7**
GeoFormer	68.0	76.8	85.4	76.7	19.9	43.8	68.4	44.0

Table 8: **Split-view of HPatches results**.

	Illumination			Viewpoint		
	3px	5px	10px	3px	5px	10px
Training data: *Oxford-Paris*						
LoFTR	77.0	85.6	92.6	40.7	54.7	70.2
ASpanFormer	74.4	84.1	91.8	46.0	58.6	71.8
DKM	31.3	36.7	41.6	30.7	38.5	46.6
GeoFormer	**86.6**	**91.6**	**95.6**	**49.3**	**62.4**	**75.9**
Training data: *MegaDepth*						
LoFTR	79.6	87.2	93.4	51.7	63.7	76.3
ASpanFormer	79.5	87.2	93.4	54.9	67.0	78.1
DKM	79.9	87.5	93.6	**62.1**	**73.4**	**83.6**
GeoFormer	**89.1**	**93.2**	**96.4**	55.7	67.4	79.3

Table 9: **Performance of visual localization on InLoc**. Training data: MegaDepth.

Method	DUC1	DUC2
	(0.25m, 10°) / (0.5m, 10°) / (1m, 10°)	
LoFTR	46.5 / 70.2 / 81.3	48.9 / 67.9 / 80.9
GeoFromer	40.9 / 65.7 / 75.8	45.0 / 62.6 / 77.1

[2] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperPoint: Self-supervised interest point detection and description. In *CVPRW*, 2018. 1

[3] Johan Edstedt, Ioannis Athanasiadis, Mårten Wadenbäck, and Michael Felsberg. Dkm: Dense kernelized feature matching for geometry estimation. In *CVPR*, 2023. 1

[4] Bumsub Ham, Minsu Cho, Cordelia Schmid, and Jean Ponce. Proposal flow: Semantic correspondences from object proposals. *IEEE transactions on pattern analysis and machine intelligence*, 40(7):1711–1725, 2017. 1

[5] Mingbo Hong, Yuhang Lu, Nianjin Ye, Chunyu Lin, Qijun Zhao, and Shuaicheng Liu. Unsupervised homography estimation with coplanarity-aware GAN. In *CVPR*, 2022. 1

[6] Zhengqi Li and Noah Snavely. Megadepth: Learning single-view depth prediction from internet photos. In *CVPR*, 2018. 1

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, 2014. 1

[8] Shuaicheng Liu, Nianjin Ye, Chuan Wang, Kunming Luo, Jue Wang, and Jian Sun. Content-aware unsupervised deep homography estimation and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (01):1–1, 2022. 1

[9] Jerome Revaud, Philippe Weinzaepfel, César Roberto de Souza, and Martin Humenberger. R2D2: Repeatable and reliable detector and descriptor. In *NeurIPS*, 2019. 1

[10] Ignacio Rocco, Mircea Cimpoi, Relja Arandjelović, Akihiko Torii, Tomas Pajdla, and Josef Sivic. NCNet: Neighbourhood consensus networks for estimating image correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(2):1020–1034, 2022. 1

[11] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *CVPR*, 2020. 1

[12] Torsten Sattler, Will Maddern, Carl Toft, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, Masatoshi Okutomi, Marc Pollefeys, Josef Sivic, et al. Benchmarking 6dof outdoor visual localization in changing conditions. In *CVPR*, 2018. 1

[13] Torsten Sattler, Tobias Weyand, Bastian Leibe, and Leif Kobbelt. Image retrieval for image-based localization revisited. In *BMVC*, 2012. 1

[14] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. LoFTR: Detector-free local feature matching with transformers. In *CVPR*, 2021. 1

[15] Erik Wijmans and Yasutaka Furukawa. Exploiting 2d floorplan for building-scale panorama rgbd alignment. In *CVPR*, 2017. 1

[16] Nianjin Ye, Chuan Wang, Haoqiang Fan, and Shuaicheng Liu. Motion basis learning for unsupervised deep homography estimation with subspace projection. In *ICCV*, 2021. 1