FC-GNN: Recovering Reliable and Accurate Correspondences from Interferences

Supplementary Material

In the following pages, we present additional experimental details, results across more matching pipelines, and more qualitative examples.

6. Experimental Details

Training data. We present a more detailed explanation of the data generation process. As described in the paper, the training data is generated by progressively introducing noise to the ground-truth matching data. Initially, a probability value, denoted as P, is defined to indicate the likelihood of adding outlier noise to the data. The inlier noise is then added to the remaining portion of the data (it is unnecessary and inappropriate to introduce both types of noise simultaneously to one match). The probability value P is uniformly distributed between 0 and 0.95. For the matches, outlier noise is introduced by randomly selecting points from the image matching space to replace them. On the other hand, inlier noise is represented by Gaussian noise, which is applied solely to the queried point. The characteristics of the inlier noise are controlled by a set of parameters. Specifically, for a given match m, the two essential parameters that determine the bias are the radius R and the angle α. The radius follows an absolute value Gaussian distribution with a standard deviation of δ, where δ is uniformly distributed between 0 and 10. The angle α adheres to a standard distribution ranging from 0 to 2π. Then, the offset of a queried point in two directions can be expressed as:

$$(x_n, y_n) = (R \cdot \sin(\alpha), R \cdot \cos(\alpha)) \tag{13}$$

It is noteworthy that, even after the addition of noise, the mean deviation for the "inliers" within the matching set remains zero. To prevent the neural network from learning this bias, a small bias (e_1, e_2) is uniformly added to these points. The biases in the two directions are drawn from a standard normal distribution. Following the incorporation of these noises, the correspondence set is compared to the original set, and matches with an error radius exceeding 8 pixels are classified as incorrect matches.

Test settings. In matching pipelines, for SIFT [27] + MNN, we use the implementation in OpenCV, the parameters of it are set to default. For SP [10] + MNN, we use the default settings in its released code. For SP [10] + SG [36], the the nms_radius is set to 3, and the $max\_keypoints$ is set to 4096 to make its matching performance better. In matching refiners, for OANet [49], we use the method it reported in the paper to remove the outliers (tanh + ReLU). For Patch2Pix [52], we adopt the settings in its paper, setting $c = 0.9$ for the image matching task and $c = 0.25$ for geometric estimation tasks.

7. Additional Results on Homography Estimation

As shown in Tab. 1, we report the combined results on more matching pipelines for geometric task on HPatches [1]. For interpretable experimental results, we crop images to a $4:3$ aspect ratio and scale them to 640×480. Homography matrix $\hat{\mathcal{H}}$ is estimated using RANSAC [15] as the robust estimator, with default parameters. We calculate reprojection errors for the four corners using both the estimated $\hat{\mathcal{H}}$ and ground-truth $\mathcal{H}$. AUC values of corner errors [40] are reported with thresholds of 3, 5, and 10. We report the results for illumination, viewpoint, and overall, respectively.

8. Additional results of Pose Estimation

As shown in Tab. 2, Tab. 3, we report the combined results on more matching pipelines for pose estimation [8, 22]. For outdoor estimation, we select two scenes, "Sacre Coeur" and "St. Peter's Square" from MegaDepth [22] dataset and use 1500 pairs of images sampled by [40]. These scenes are excluded during our training. We estimate relative pose using RANSAC and evaluate the pose error's AUC with thresholds of 5°, 10°, and 20° following [6, 36, 40]. Unlike [6, 40], we do not apply image scaling for higher precision. For indoor pose estimation, we use 1500 test image pairs selected from ScanNet [8] dataset by [40]. We resize the image size to 640×480. Similar to Ourdoor, we report the AUC of the pose error for thresholds 5°, 10°, and 20° respectively.

9. Additional Qualitative Results

The visualized experimental results are provided in Fig. 1. For better observation, the unified size of the image was set to 640×480, the threshold of Patch2Pix [52] is set to 0.5, and the threshold of FC-GNN is set to 0.9.

Matcher	Refiner	Illumination	Viewpoint	Overall	#Matches
		AUC (%, @3, 5, 10px)			
SIFT [27] + MNN	Origin	69.4 / 78.3 / 85.9	50.3 / 62.4 / 75.3	59.6 / 70.1 / 80.4	0.84k
	OANet [49]	68.3 / 77.9 / 85.7	41.9 / 56.8 / 72.8	54.8 / 67.1 / 79.1	0.39k
	Patch2Pix [52]	70.3 / 79.6 / 87.6	40.1 / 53.8 / 70.0	54.8 / 66.4 / 78.6	0.60k
	FC-GNN	**72.4 / 81.3 / 88.8**	**51.8 / 63.9 / 77.2**	**61.9 / 72.4 / 82.8**	0.56k
ORB [35] + MNN	Origin	47.2 / 54.0 / 61.6	28.7 / 39.4 / 51.4	37.7 / 46.5 / 56.4	1.24k
	OANet [49]	53.5 / 64.2 / 76.0	26.1 / 39.4 / 57.0	39.5 / 51.5 / 66.3	0.41k
	Patch2Pix [52]	58.0 / 69.4 / 79.7	30.8 / 44.6 / 61.8	44.1 / 56.7 / 70.5	0.65k
	FC-GNN	**60.4 / 70.6 / 79.1**	**40.8 / 53.6 / 67.1**	**50.3 / 61.9 / 73.0**	0.65k
SURF [3] + MNN	Origin	60.8 / 71.5 / 81.1	38.2 / 51.1 / 65.8	49.2 / 61.0 / 73.2	0.83k
	OANet [49]	60.9 / 73.0 / 83.8	32.5 / 47.2 / 64.6	46.4 / 59.8 / 74.0	0.36k
	Patch2Pix [52]	**69.8 / 79.9 / 88.8**	38.3 / 52.9 / 69.8	53.7 / 66.1 / 79.0	0.57k
	FC-GNN	69.1 / 79.1 / 87.8	**47.3 / 60.1 / 74.4**	**57.9 / 69.3 / 80.9**	0.51k
D2Net [12] + MNN	Origin	30.7 / 52.7 / 74.8	7.4 / 19.6 / 42.4	18.8 / 35.8 / 58.2	1.14k
	OANet [49]	17.6 / 35.0 / 62.1	2.8 / 10.1 / 28.5	10.0 / 22.3 / 44.9	0.50k
	Patch2Pix [52]	70.5 / 81.7 / 90.7	38.9 / 52.8 / 68.9	54.3 / 66.9 / 79.5	1.06k
	FC-GNN	**75.6 / 84.7 / 92.2**	**41.7 / 55.1 / 69.8**	**58.3 / 69.6 / 80.7**	1.06k
R2D2 [31] + MNN	Origin	64.8 / 78.0 / 88.6	36.6 / 49.8 / 65.7	50.4 / 63.6 / 76.9	1.59k
	OANet [49]	54.2 / 69.5 / 84.3	24.7 / 37.4 / 55.5	39.1 / 53.1 / 69.6	0.94k
	Patch2Pix [52]	74.8 / 84.0 / 91.7	38.5 / 51.3 / 66.0	56.2 / 67.3 / 78.5	1.54k
	FC-GNN	**76.1 / 85.1 / 92.4**	**44.1 / 56.1 / 69.7**	**59.7 / 70.3 / 80.8**	1.54k
SP [10] + MNN	Origin	61.4 / 74.8 / 87.0	38.0 / 53.5 / 69.8	49.4 / 63.9 / 78.2	0.55k
	OANet [49]	46.6 / 64.0 / 80.7	24.2 / 38.2 / 56.7	35.1 / 50.8 / 68.4	0.25k
	Patch2Pix [52]	72.9 / 83.0 / 91.4	41.4 / 55.0 / 69.8	56.8 / 68.7 / 80.3	0.51k
	FC-GNN	**75.2 / 84.4 / 92.0**	**48.9 / 61.3 / 74.9**	**61.8 / 72.6 / 83.3**	0.49k
SP [10] + SG [36]	Origin	62.6 / 76.4 / 88.1	45.7 / 61.3 / 76.8	54.0 / 68.6 / 82.3	0.61k
	OANet [49]	37.4 / 56.5 / 76.2	20.1 / 35.6 / 56.8	28.5 / 45.8 / 66.3	0.14k
	Patch2Pix [52]	77.3 / 85.9 / 91.2	42.4 / 56.2 / 72.6	57.0 / 69.2 / 81.7	0.58k
	FC-GNN	**76.9 / 85.9 / 92.9**	**57.2 / 69.0 / 81.3**	**67.0 / 77.2 / 86.9**	0.60k
DRC-Net [21]	Origin	**95.8 / 96.7 / 98.1**	13.1 / 29.4 / 52.2	53.7 / 62.4 / 74.7	1.94k
	OANet [49]	95.2 / 95.9 / 97.6	3.5 / 10.3 / 26.0	48.5 / 52.3 / 61.1	0.47k
	Patch2Pix [52]	75.2 / 84.5 / 92.0	**33.3 / 47.6 / 65.4**	53.8 / 65.6 / 78.5	1.87k
	FC-GNN	82.6 / 89.0 / 94.3	32.3 / **47.6** / 64.8	**56.9 / 67.9 / 79.3**	1.93k
LoFTR [40]	Origin	**80.4 / 87.5 / 93.5**	48.7 / 60.1 / 74.5	64.2 / 74.0 / 83.8	2.68k
	OANet [49]	70.7 / 81.3 / 90.2	31.4 / 45.6 / 63.5	50.6 / 63.1 / 76.5	0.91k
	Patch2Pix [52]	70.8 / 81.6 / 90.6	40.7 / 55.6 / 71.4	55.4 / 68.3 / 80.8	2.54k
	FC-GNN	80.2 / **87.5 / 93.5**	**52.5 / 64.5 / 76.9**	**66.1 / 75.8 / 84.9**	2.66k
SIFT [27] + LG [24]	Origin	69.7 / 80.4 / 89.7	53.8 / 66.5 / 79.5	61.5 / 73.3 / 84.4	1.45k
	OANet [49]	50.0 / 66.4 / 81.8	33.8 / 47.8 / 65.2	41.7 / 56.9 / 73.3	0.34k
	Patch2Pix [52]	70.2 / 81.2 / 90.3	39.8 / 54.8 / 72.4	54.6 / 67.7 / 81.1	1.39k
	FC-GNN	**76.6 / 85.3 / 92.3**	**56.1 / 68.7 / 81.2**	**66.1 / 76.8 / 86.6**	1.44k
DISK [42] + LG [24]	Origin	65.0 / 77.2 / 88.4	48.0 / 60.8 / 74.5	56.3 / 68.8 / 81.2	2.26k
	OANet [49]	42.8 / 59.5 / 77.2	26.5 / 42.0 / 60.7	34.4 / 50.5 / 68.7	0.52k
	Patch2Pix [52]	64.4 / 76.5 / 87.6	36.3 / 51.2 / 68.1	50.0 / 63.6 / 77.6	2.18k
	FC-GNN	**74.9 / 83.8 / 91.5**	**50.7 / 63.1 / 75.6**	**62.5 / 73.2 / 83.3**	2.26k
ALIKED [50] + LG [24]	Origin	71.7 / 82.2 / 90.8	52.5 / 65.5 / 78.6	61.9 / 73.6 / 84.5	1.14k
	OANet [49]	50.0 / 66.0 / 81.6	31.9 / 46.2 / 63.7	40.7 / 55.8 / 72.4	0.27k
	Patch2Pix [52]	70.2 / 81.3 / 90.4	38.7 / 53.6 / 70.4	54.1 / 67.2 / 80.1	1.11k
	FC-GNN	**75.7 / 84.6 / 91.9**	**56.2 / 68.2 / 80.0**	**65.7 / 76.2 / 85.7**	1.14k
ASpanFormer [6]	Origin	**80.0 / 87.3 / 93.4**	49.2 / 62.1 / 75.4	64.3 / 74.4 / 84.2	2.76k
	OANet [49]	66.8 / 78.9 / 88.8	32.7 / 46.4 / 63.9	49.4 / 62.3 / 76.1	0.68k
	Patch2Pix [52]	69.4 / 80.8 / 90.3	39.9 / 54.1 / 70.0	54.4 / 67.2 / 79.9	2.56k
	FC-GNN	79.4 / 87.2 / **93.4**	**53.4 / 65.5 / 78.1**	**66.1 / 76.1 / 85.5**	2.75k

Table 1. **Homography estimation on HPatches [1]**. The AUC of the corner error in percentage is reported. We mark the best results in **bold**.

Matcher	Refiner	Pose estimation AUC @5°	@10°	@20°
SIFT [27] + MNN	Origin	16.67	28.54	42.75
	OANet [49]	40.28	57.07	71.00
	Patch2Pix [52]	33.54	48.66	62.07
	FC-GNN	**44.57**	**60.32**	**72.78**
SURF [3] + MNN	Origin	10.71	21.69	36.14
	OANet [49]	32.84	50.62	66.34
	Patch2Pix [52]	31.16	47.05	61.66
	FC-GNN	**40.08**	**57.26**	**71.28**
ORB [35] + MNN	Origin	2.70	6.50	13.48
	OANet [49]	13.72	23.83	36.30
	Patch2Pix [52]	15.77	26.80	39.50
	FC-GNN	**19.17**	**30.48**	**42.30**
D2Net [12] + MNN	Origin	21.88	37.50	53.49
	OANet [49]	15.76	31.06	48.78
	Patch2Pix [52]	42.40	57.95	70.42
	FC-GNN	**44.05**	**59.10**	**70.73**
R2D2 [31] + MNN	Origin	44.04	61.50	74.77
	OANet [49]	33.05	51.69	67.90
	Patch2Pix [52]	42.20	58.44	71.58
	FC-GNN	**48.87**	**65.70**	**78.01**
SP [10] + MNN	Origin	30.00	45.24	59.29
	OANet [49]	31.59	49.30	64.32
	Patch2Pix [52]	39.29	54.83	67.27
	FC-GNN	**45.58**	**60.92**	**72.19**
SP [10] + SG [36]	Origin	49.13	66.16	79.23
	OANet [49]	23.40	40.31	58.36
	Patch2Pix [52]	47.32	63.98	77.23
	FC-GNN	**54.67**	**71.03**	**82.65**
SIFT [27] + LG [24]	Origin	50.51	67.33	80.45
	OANet [49]	23.88	41.62	60.22
	Patch2Pix [52]	45.10	61.36	74.40
	FC-GNN	**52.39**	**69.57**	**81.88**
DISK [42] + LG [24]	Origin	45.43	63.04	76.92
	OANet [49]	23.86	41.01	59.15
	Patch2Pix [52]	42.05	59.05	72.84
	FC-GNN	**50.87**	**67.86**	**80.50**
ALIKED [50] + LG [24]	Origin	47.51	65.25	78.85
	OANet [49]	24.95	41.65	58.71
	Patch2Pix [52]	43.27	60.18	73.56
	FC-GNN	**51.00**	**68.24**	**80.78**
ASpanFormer [6]	Origin	55.36	71.31	82.90
	OANet [49]	37.62	56.74	72.96
	Patch2Pix [52]	48.17	64.46	77.25
	FC-GNN	**56.64**	**72.43**	**83.76**

Table 2. **Outdoor pose estimation.** The AUC of the pose error in percentage is reported. We mark the best results in **bold**.

Matcher	Refiner	Pose estimation AUC @5°	@10°	@20°
SIFT [27] + MNN	Origin	4.26	10.10	18.11
	OANet [49]	5.88	13.58	23.22
	Patch2Pix [52]	6.09	14.07	24.62
	FC-GNN	**7.88**	**16.87**	**27.59**
SURF [3] + MNN	Origin	4.02	10.71	20.46
	OANet [49]	7.14	16.68	29.54
	Patch2Pix [52]	8.34	18.98	32.48
	FC-GNN	**10.67**	**22.30**	**35.46**
ORB [35] + MNN	Origin	1.54	4.63	9.92
	OANet [49]	4.22	10.05	18.56
	Patch2Pix [52]	5.42	12.53	**22.12**
	FC-GNN	**5.47**	**12.56**	21.66
D2Net [12] + MNN	Origin	3.62	10.89	22.82
	OANet [49]	2.37	6.69	14.30
	Patch2Pix [52]	**11.01**	**23.18**	**36.47**
	FC-GNN	10.34	21.91	35.52
R2D2 [31] + MNN	Origin	7.83	17.09	28.72
	OANet [49]	5.82	14.05	24.34
	Patch2Pix [52]	8.61	19.24	31.14
	FC-GNN	**10.47**	**22.07**	**34.82**
SP [10] + MNN	Origin	8.79	19.51	32.51
	OANet [49]	7.15	16.96	29.48
	Patch2Pix [52]	12.01	25.83	40.91
	FC-GNN	**14.47**	**29.45**	**44.50**
SP [10] + SG [36]	Origin	15.31	31.64	48.00
	OANet [49]	3.56	9.66	19.83
	Patch2Pix [52]	15.33	31.74	48.10
	FC-GNN	**18.46**	**36.47**	**52.98**
SIFT [27] + LG [24]	Origin	15.08	30.52	46.17
	OANet [49]	5.67	14.83	27.45
	Patch2Pix [52]	13.72	28.62	43.96
	FC-GNN	**17.34**	**33.96**	**49.66**
DISK [42] + LG [24]	Origin	12.74	24.80	38.32
	OANet [49]	5.16	12.32	22.59
	Patch2Pix [52]	11.69	24.26	37.94
	FC-GNN	**13.59**	**26.81**	**40.74**
ALIKED [50] + LG [24]	Origin	14.63	29.21	44.03
	OANet [49]	5.10	11.70	21.79
	Patch2Pix [52]	13.31	27.31	41.94
	FC-GNN	**16.79**	**31.83**	**46.68**
ASpanFormer [6]	Origin	25.69	45.85	63.31
	OANet [49]	9.57	22.06	37.69
	Patch2Pix [52]	16.45	32.98	48.73
	FC-GNN	**26.01**	**46.43**	**63.90**

Table 3. **Indoor pose estimation.** The AUC of the pose error in percentage is reported. We mark the best results in **bold**.

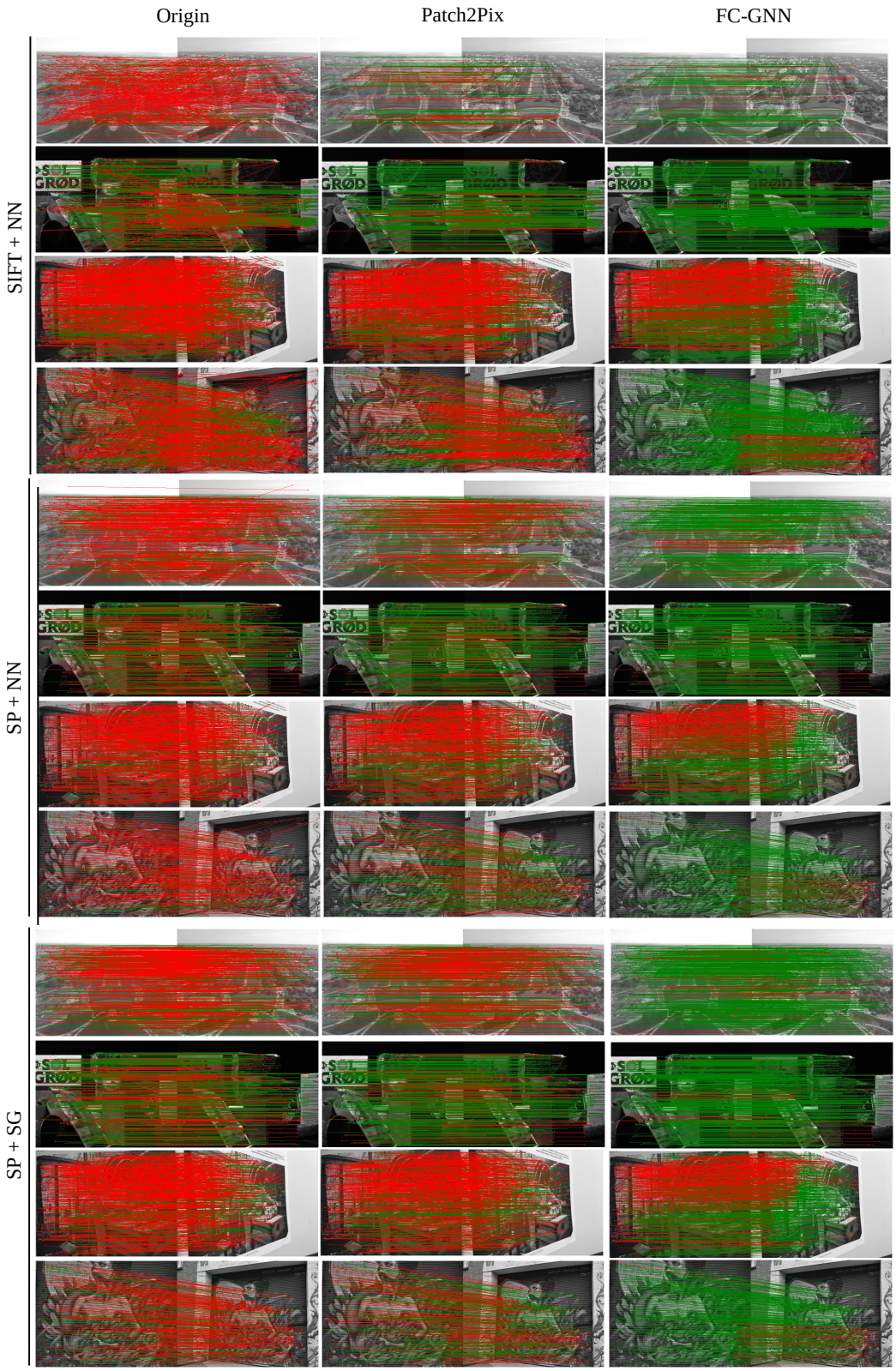

Figure 1. **Qualitative image matches on HPatches [1].** We mark matches with an error ≤ 1 pixel as green, and the rest as red. It can be seen that FC-GNN greatly improves the accuracy of matching and effectively filtering out outliers.

References

[1] Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *CVPR*, 2017. 6, 7, 8, 1, 2, 4

[2] Axel Barroso-Laguna, Edgar Riba, Daniel Ponsa, and Krystian Mikolajczyk. Key. net: Keypoint detection by handcrafted and learned cnn filters. In *CVPR*, 2019. 2

[3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *ECCV*, 2006. 2, 3

[4] Eric Brachmann and Carsten Rother. Neural-guided ransac: Learning where to sample model hypotheses. In *CVPR*, 2019. 1, 2

[5] Hongkai Chen, Zixin Luo, Jiahui Zhang, Lei Zhou, Xuyang Bai, Zeyu Hu, Chiew-Lan Tai, and Long Quan. Learning to match features with seeded graph matching network. In *ICCV*, 2021. 1, 2, 4

[6] 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. 2, 5, 6, 7, 1, 3

[7] Xiangxiang Chu, Bo Zhang, Zhi Tian, Xiaolin Wei, and Huaxia Xia. Do we really need explicit position encodings for vision transformers. *arXiv preprint arXiv:2102.10882*, 3 (8), 2021. 4

[8] Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *CVPR*, 2017. 7, 1

[9] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Deep image homography estimation. *arXiv preprint arXiv:1606.03798*, 2016. 1

[10] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *CVPR*, 2018. 2, 5, 6, 7, 8, 1, 3

[11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 4

[12] Mihai Dusmanu, Ignacio Rocco, Tomas Pajdla, Marc Pollefeys, Josef Sivic, Akihiko Torii, and Torsten Sattler. D2-net: A trainable cnn for joint description and detection of local features. In *CVPR*, 2019. 2, 3

[13] Mihai Dusmanu, Johannes L Schönberger, and Marc Pollefeys. Multi-view optimization of local feature geometry. In *ECCV*, 2020. 2

[14] Ufuk Efe, Kutalmis Gokalp Ince, and Aydin Alatan. Dfm: A performance baseline for deep feature matching. In *CVPR*, 2021. 2

[15] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981. 2, 7, 1

[16] Hugo Germain, Guillaume Bourmaud, and Vincent Lepetit. S2dnet: Learning accurate correspondences for sparse-to-dense feature matching. *arXiv preprint arXiv:2004.01673*, 2020. 2

[17] Pierre Gleize, Weiyao Wang, and Matt Feiszli. Silk–simple learned keypoints. *arXiv preprint arXiv:2304.06194*, 2023. 2

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 4, 5

[19] Jogendra Nath Kundu, MV Rahul, Aditya Ganeshan, and R Venkatesh Babu. Object pose estimation from monocular image using multi-view keypoint correspondence. In *ECCV*, 2018. 1

[20] Jongmin Lee, Byungjin Kim, and Minsu Cho. Self-supervised equivariant learning for oriented keypoint detection. In *CVPR*, 2022. 2

[21] Xinghui Li, Kai Han, Shuda Li, and Victor Prisacariu. Dual-resolution correspondence networks. In *NeurIPS*, 2020. 2

[22] Zhengqi Li and Noah Snavely. Megadepth: Learning single-view depth prediction from internet photos. In *CVPR*, 2018. 5, 7, 1

[23] Philipp Lindenberger, Paul-Edouard Sarlin, Viktor Larsson, and Marc Pollefeys. Pixel-perfect structure-from-motion with featuremetric refinement. In *CVPR*, 2021. 2, 5

[24] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching at light speed. *arXiv preprint arXiv:2306.13643*, 2023. 2, 3

[25] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *CVPR*, 2021. 4

[26] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 5

[27] David G Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004. 2, 5, 6, 7, 1, 3

[28] Zixin Luo, Tianwei Shen, Lei Zhou, Jiahui Zhang, Yao Yao, Shiwei Li, Tian Fang, and Long Quan. Contextdesc: Local descriptor augmentation with cross-modality context. In *CVPR*, 2019. 2

[29] Zixin Luo, Lei Zhou, Xuyang Bai, Hongkai Chen, Jiahui Zhang, Yao Yao, Shiwei Li, Tian Fang, and Long Quan. Aslfeat: Learning local features of accurate shape and localization. In *CVPR*, 2020. 2

[30] Anastasiia Mishchuk, Dmytro Mishkin, Filip Radenovic, and Jiri Matas. Working hard to know your neighbor's margins: Local descriptor learning loss. *NeurIPS*, 30, 2017.

[31] Jerome Revaud, Philippe Weinzaepfel, César De Souza, Noe Pion, Gabriela Csurka, Yohann Cabon, and Martin Humenberger. R2d2: repeatable and reliable detector and descriptor. *arXiv preprint arXiv:1906.06195*, 2019. 2, 3

[32] Jérome Revaud, Vincent Leroy, Philippe Weinzaepfel, and Boris Chidlovskii. Pump: Pyramidal and uniqueness matching priors for unsupervised learning of local descriptors. In *CVPR*, 2022. 2

[33] Ignacio Rocco, Mircea Cimpoi, Relja Arandjelović, Akihiko Torii, Tomas Pajdla, and Josef Sivic. Neighbourhood consensus networks. In *NeurIPS*, 2018. 2

[34] Ignacio Rocco, Relja Arandjelović, and Josef Sivic. Efficient neighbourhood consensus networks via submanifold sparse convolutions. In *ECCV*, 2020. 2

[35] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *ICCV*, 2011. 2, 3

[36] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *CVPR*, 2020. 1, 2, 4, 5, 6, 7, 3

[37] Nikolay Savinov, Akihito Seki, Lubor Ladicky, Torsten Sattler, and Marc Pollefeys. Quad-networks: unsupervised learning to rank for interest point detection. In *CVPR*, 2017. 2

[38] Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. *arXiv preprint arXiv:1803.02155*, 2018. 4

[39] Yan Shi, Jun-Xiong Cai, Yoli Shavit, Tai-Jiang Mu, Wensen Feng, and Kai Zhang. Clustergnn: Cluster-based coarse-to-fine graph neural network for efficient feature matching. In *CVPR*, 2022. 1, 2, 4

[40] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. Loftr: Detector-free local feature matching with transformers. In *CVPR*, 2021. 2, 7, 1

[41] Weiwei Sun, Wei Jiang, Eduard Trulls, Andrea Tagliasacchi, and Kwang Moo Yi. Acne: Attentive context normalization for robust permutation-equivariant learning. In *CVPR*, 2020. 2, 5

[42] Michał Tyszkiewicz, Pascal Fua, and Eduard Trulls. Disk: Learning local features with policy gradient. In *NeurIPS*, 2020. 2, 3

[43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017. 2, 4

[44] Qianqian Wang, Xiaowei Zhou, Bharath Hariharan, and Noah Snavely. Learning feature descriptors using camera pose supervision. In *ECCV*, 2020. 2

[45] Qing Wang, Jiaming Zhang, Kailun Yang, Kunyu Peng, and Rainer Stiefelhagen. Matchformer: Interleaving attention in transformers for feature matching. *arXiv preprint arXiv:2203.09645*, 2022. 2

[46] Haiping Wu, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, and Lei Zhang. Cvt: Introducing convolutions to vision transformers. In *CVPR*, 2021. 4

[47] Fei Xue, Ignas Budvytis, and Roberto Cipolla. Imp: Iterative matching and pose estimation with adaptive pooling. In *CVPR*, 2023. 2

[48] Kwang Moo Yi, Eduard Trulls, Yuki Ono, Vincent Lepetit, Mathieu Salzmann, and Pascal Fua. Learning to find good correspondences. In *CVPR*, 2018. 1, 2, 5

[49] Jiahui Zhang, Dawei Sun, Zixin Luo, Anbang Yao, Lei Zhou, Tianwei Shen, Yurong Chen, Long Quan, and Hongen Liao. Learning two-view correspondences and geometry using order-aware network. In *ICCV*, 2019. 1, 2, 5, 6, 7, 3

[50] Xiaoming Zhao, Xingming Wu, Weihai Chen, Peter CY Chen, Qingsong Xu, and Zhengguo Li. Aliked: A lighter keypoint and descriptor extraction network via deformable transformation. *IEEE Transactions on Instrumentation and Measurement*, 2023. 2, 3

[51] Qunjie Zhou, Torsten Sattler, Marc Pollefeys, and Laura Leal-Taixe. To learn or not to learn: Visual localization from essential matrices. In *Proceedings of IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2020. 1

[52] Qunjie Zhou, Torsten Sattler, and Laura Leal-Taixe. Patch2pix: Epipolar-guided pixel-level correspondences. In *CVPR*, 2021. 2, 4, 5, 6, 7, 1, 3