

LoMa: Local Feature Matching Revisited

Supplementary Material

A Additional Experiments

A.1 Performance using Different Detectors

All the results in the main paper use the DaD [19] detector. We investigate the performance of LoMa using different detectors in Tab. 6. For fair comparison, we retrain the descriptor with an ensemble of detectors and use this for all the subsequent comparisons. To create an ensemble [68], we uniformly sample keypoints from DeDoDe v2 [17], DISK [61], ALIKED [76], and DaD [19] during training. We then train separate matchers (one for each detector) and one ensemble matcher that is jointly trained with keypoints randomly sampled from all detectors. We find DaD to be overall the strongest detector, regardless of setting. Generally, slightly better performance is achieved by specializing the matcher on one detector.

Table 6: Detector ablation. Performance (mAA@10px on the validation set of Hard-Match) comparison between training with a single detector and randomly sampling multiple detectors (ensemble).

Method	Ensemble Training	Single Detector
DeDoDe v2 [17]	47.0	47.3
DISK [61]	44.3	45.3
ALIKED [76]	48.9	49.0
DaD [19]	51.0	51.3

A.2 Dependence on the Number of Keypoints

Throughout the main paper, we evaluate with $N = 4096$ keypoints. We study the performance for fewer keypoints in Tab. 7. The performance difference between 2048 and 4096 keypoints is negligible for other sparse matchers, but LoMa benefits slightly from increasing the number of keypoints beyond 2048. Performance degrades significantly below 2048 keypoints.

A.3 HPatches

We evaluate on HPatches [5] following the LoFTR [56] protocol. The dataset contains planar scenes with homographies. We report the results in Tab. 8. The lightweight LoMa-B¹²⁸ achieves the highest score.

Table 7: Varying max number of keypoints. We report the AUC@20 for different number of maximum keypoints.

Method	MegaDepth-1500				ScanNet-1500			
	512	1024	2048	4096	512	1024	2048	4096
SP+SG	70.9	75.4	76.5	76.4	44.5	48.1	49.0	49.1
SP+LG	70.6	74.8	76.4	76.4	43.9	47.6	48.6	48.9
LoMa-B (ours)	80.3	82.7	83.1	83.6	58.3	63.3	66.4	68.2

Table 8: HPatches. Performance on HPatches [5].

Method	HPatches		
	AUC@→	@3px	@5px @10px
SP+SG	64.4	75.6	86.0
SP+LG	64.2	75.5	85.5
DISK+LG	60.4	72.4	83.5
ALIKED+LG	66.2	76.9	86.3
LoMa-B ¹²⁸ (ours)	67.2	78.1	87.8
LoMa-B (ours)	66.5	77.5	87.3
LoMa-L (ours)	66.5	77.8	87.5
LoMa-G (ours)	66.4	77.5	87.3

A.4 SatAst

We evaluate astronaut to satellite matching on SatAst [20]. The dataset features large in-plane rotations and scale changes, making it difficult for most matchers. We report the results in Tab. 4, where we find that, while beating other sparse matchers, LoMa struggles with rotations.

A.5 Oxford Day-and-Night

In the main paper, we report the median performance on the night queries of Oxford Day-and-Night [71]. We use the HLoc [51] pipeline with NetVLAD-50 [2]. The benchmark contains four outdoor scenes (Bodleian Library, H.B. Allen Centre, Keble College, Observatory Quarter) and one indoor scene (Robotics Institute). In Tab. 10, we report the results per scene.

A.6 WxBS

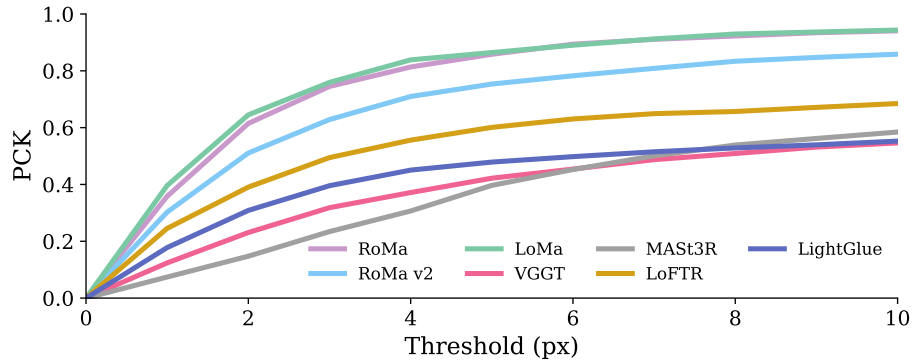
To better understand the relative performance of the matchers, we report the accuracy (PCK) at different pixel thresholds in Fig. 9.

Table 9: SatAst. Astronaut to satellite matching on SatAst [20].

Method	SatAst
AUC@ $\rightarrow$	@10px
SP+SG	19.8
SP+LG	12.8
DISK+LG	0.0
ALIKED+LG	12.1
LoMa-B (ours)	18.8
LoMa-L (ours)	21.3
LoMa-G (ours)	22.9

Table 10: Oxford Day-and-Night. Full results for night queries. Reporting percentage of correctly localized test images within (0.25m, 2°) / (0.5m, 5°) / (1m, 10°)

	Bodleian Library	H.B. Allen Centre	Keble College	Observatory Quarter	Robotics Institute
SP+SG	21.6/26.5/30.8	44.3/57.5/64.1	10.7/13.6/17.3	48.1/54.4/58.0	71.1/73.6/74.5
SP+LG	20.5/25.3/28.8	43.4/54.1/61.5	10.0/14.2/18.3	47.9/53.5/57.7	70.1/71.8/73.0
DISK+LG	14.8/17.5/20.1	9.6/11.4/14.7	0.5/0.8/1.1	16.8/20.4/22.9	53.2/57.5/60.5
ALIKED+LG	22.2/27.3/31.0	42.8/55.5/62.6	10.3/13.5/18.3	45.1/53.9/58.9	57.2/61.4/63.6
LoMa-B ₁₂₈ (ours)	26.1/31.8/36.1	60.1/71.1/76.6	14.2/17.8/22.0	54.8/62.2/67.1	73.3/76.1/77.0
LoMa-B (ours)	24.1/28.9/32.5	57.5/69.7/73.7	15.3/21.0/25.2	54.7/62.4/66.2	70.9/74.1/74.7
LoMa-L (ours)	25.1/30.4/34.3	62.1/71.9/76.0	15.4/21.1/26.0	56.0/64.6/69.2	71.5/74.9/76.0
LoMa-G (ours)	25.1/30.4/35.0	61.9/ 73.7/78.2	15.8/21.4/27.5	58.9/66.0/69.7	72.2/75.2/76.5

**Fig. 9: WxBS accuracy at different thresholds.**

A.7 Scaling the Descriptor

As shown in Fig. 10a, the HardMatch validation loss of the descriptor saturates at around 50K steps and then slowly increases. Thus, we limit the descriptor training to only 50K steps (compared to 250K for the matcher).

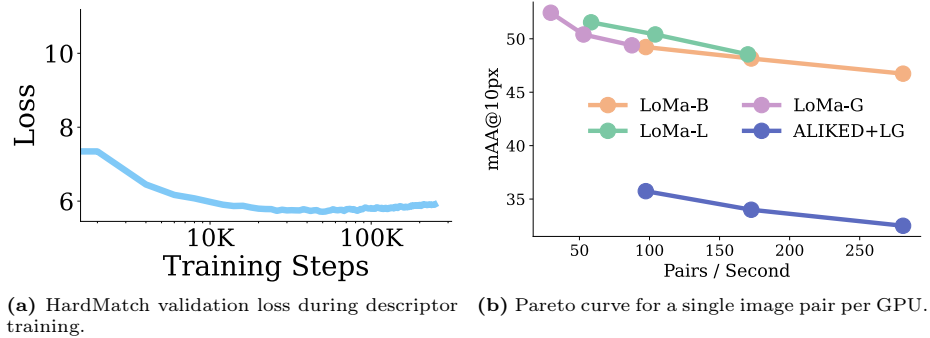


Fig. 10: Descriptor scaling (a) and pareto curve for batch size of 1 (b).

A.8 Inference Speed for a Single Image Pair

In the main paper, we evaluate the speed using a batch size of 16. In many applications, inference will run with a single image pair at a time (batch size of 1). In Fig. 10b, we show the speed for our different model sizes for different stopping layers $L = \{3, 5, 9\}$.

A.9 Map-free Visual Relocalization Benchmark

In Tab. 11, we compare to additional matchers on the validation set (using DA3 depths for consistency). We also evaluated LoMa-G on the official benchmark, achieving an AUC and precision (VCRE<90px) of 92.5 and 75.7, respectively.

Table 11: Additional Map-free Baselines. TBD

Metric	MASt3R	LoFTR	RoMa	LoMa-G
Prec. $\uparrow$	73.4	39.7	59.7	68.9
AUC $\uparrow$	91.6	73.5	84.4	90.3

B Details on Evaluation

For all LoMa evaluations, we use an internal resolution of 784×784 , $N = 4096$ DaD keypoints, and $L = 9$ layers.

B.1 Relative Pose Estimation

The evaluation protocol follows from LoFTR [56] and is also used in *e.g.* RoMa [21] and RoMa v2 [20]. We use a RANSAC pixel threshold of $\tau = 0.5$. We use the standard AUC metric. The AUC metric evaluates the error of the estimated Essential matrix relative to the ground truth. For each image pair, the error is defined as the maximum of the rotational and translational errors. Since metric scale is unavailable, the translational error is measured using the cosine of the angular difference. The recall at a threshold τ is defined as the fraction of pairs whose error is below τ . The metric $\text{AUC@}\tau^\circ$ is computed as the normalized integral of the recall curve with respect to the threshold from 0 to τ , divided by τ . In practice, this integral is approximated using the trapezoidal rule over the set of errors produced by the method on the dataset.

B.2 Image Matching Challenge 2022

We run the official Kaggle competition and use 200K iterations of MAGSAC [6, 7] with an inlier threshold of $\tau = 0.2$. We report the mean average accuracy (mAA). The metric evaluates the estimated Fundamental matrix against the hidden ground truth using rotational error (in degrees) and translational error (in meters). A pose is considered correct if both errors fall below specified thresholds. This is evaluated over ten uniformly spaced threshold pairs. The mAA is the average accuracy across all thresholds and images, balanced across scenes.

C HardMatch

C.1 Further Details on Evaluation

Following WxBS [42], our main method for evaluating the hand-labeled correspondences in HardMatch is through the estimation of a Fundamental matrix. Specifically, each method finds matches between the two images and robustly estimates the fundamental matrix using the OpenCV implementation of MAGSAC [6, 7] with an inlier threshold of $\tau = 0.25$ pixels. We compute the percent of keypoints (PCK) in the ground truth correspondences consistent with the estimated Fundamental matrix for epipolar pixel thresholds going from 0 to 20. We compute the pixel errors at a resolution of 640×640 and do not evaluate on the approximately 20 pairs we label dynamic.

C.2 Correspondence Evaluation

An alternative methodology involves directly matching the ground truth keypoints. This has the benefit of working even for image pairs where a Fundamental matrix is not well defined, *e.g.* dynamic scenes and non-perspective cameras. For dense matchers, we sample the warp at the ground truth keypoint locations in I^A and find the pixel error to the ground truth correspondences in I^B . For sparse matchers, the most straightforward way is to append the ground truth keypoints to the detected keypoints in I^A and record the pixel error between the estimated and true matches in I^B . As illustrated in Tab. 12, LoMa performs the best also on this evaluation.

Table 12: Correspondence Evaluation. HardMatch evaluation where, for sparse matchers, ground truth correspondences are appended to detected keypoints.

Method	HardMatch			
	5px	10px	15px	20px
<i>Dense Matchers</i>				
RoMa	52.9	60.9	64.1	66.0
UFM	35.2	47.1	53.4	56.7
RoMa v2	53.2	64.5	70.4	73.3
<i>Sparse Matchers, 2048 keypoints</i>				
SP+SG	37.6	39.3	40.2	40.6
SP+LG	39.3	45.1	47.8	49.5
LoMa-B (ours)	64.3	71.4	74.5	76.0
LoMa-L (ours)	65.8	72.7	76.0	77.4
LoMa-G (ours)	68.0	74.0	76.8	78.2

C.3 Dataset statistics

Images in the HardMatch dataset date between the early 20th century until now (*cf.* Fig. 11a) and has a global geographic footprint (*cf.* Fig. 11b). Most of the images are taken at the start of the 21st century in Europe.

C.4 Qualitative Pairs with Matches

In Fig. 12 we illustrate some representative examples for challenging groups in HardMatch. In Fig. 13 we display a random collection of pairs and the matches detected by LoMa-G (inliers during Fundamental matrix estimation with $\tau = 5$ are colored green).

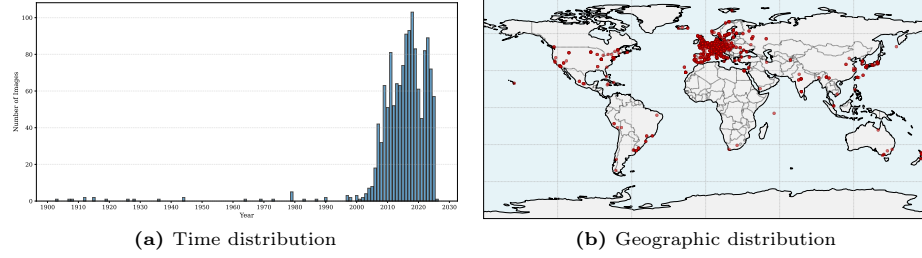


Fig. 11: HardMatch statistics. The dataset consists of images taken from all over the world and from over a century apart. The highest concentration is geographically in Europe and temporally in the 21st century.

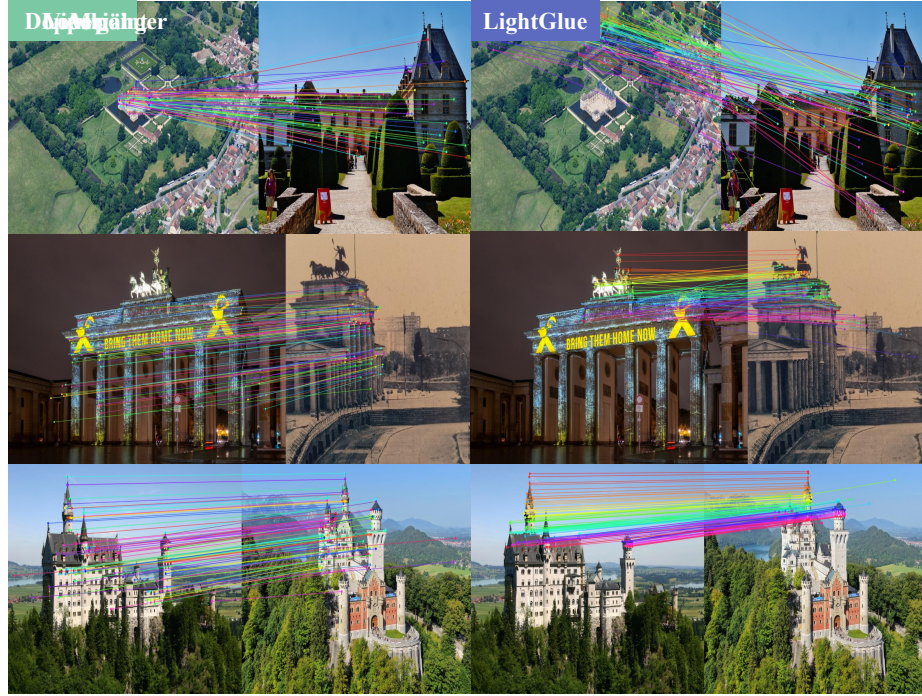


Fig. 12: Hard groups of HardMatch. For hard Doppelgängers in HardMatch, all the matchers fail.



Fig. 13: LoMa-G matches from HardMatch. Inliers at 5px threshold for MAGSAC [6,7] colored green while outliers are colored red.

C.5 Results by Category and Group

The HardMatch dataset is sourced from 100 Wikimedia Commons categories. We list the categories of the test dataset in Fig. 14. We also report the detailed performance breakdown of different groups in Tab. 14.

Table 14: Detailed HardMatch Performance. Performance (mAA@10px) on different groupings of HardMatch.

Method	Aerial	Celestial	Doppelgänger	Drawing	Illumination	Nature	Seasonal	Temporal	Viewpoint
<i>Feedforward Reconstruction</i>									
MASt3R [33]	13.0	20.2	25.4	29.1	32.3	39.5	29.7	32.3	18.2
VGGT [66]	12.9	8.7	14.8	17.9	32.2	27.7	30.6	32.0	15.6
<i>Dense Matchers</i>									
LoFTR [56]	12.9	25.9	19.3	21.3	33.5	39.5	34.3	38.6	10.5
RoMa [21]	27.2	26.1	28.7	41.2	50.0	54.2	51.3	55.0	20.8
UFM [75]	14.4	30.7	22.4	30.4	32.0	33.0	40.3	41.7	15.0
RoMa v2 [20]	28.7	28.3	34.1	34.3	49.1	54.4	46.5	50.5	28.6
<i>Sparse Matchers, 4096 Keypoints</i>									
SP+SG [14] [52]	16.7	28.1	23.5	27.2	34.5	41.5	42.2	40.8	8.6
SP+LG [14] [36]	12.9	25.1	22.1	26.8	32.8	40.0	37.2	39.2	8.3
DISK+LG [36] [61]	12.0	6.3	16.8	20.0	25.3	30.7	29.2	36.9	9.6
ALIKED+LG [36] [76]	14.0	16.4	21.5	26.0	34.2	40.9	39.1	41.6	10.5
LoMa-B (ours)	31.6	28.1	30.3	41.0	51.7	54.4	54.6	55.6	26.0
LoMa-L (ours)	35.6	35.8	35.9	43.0	54.5	55.2	56.4	57.6	30.3
LoMa-G (ours)	36.9	35.7	36.3	40.4	55.0	56.0	58.8	59.4	30.9

D Progressive Match Refinement

We qualitatively examine the detected matches at different stopping layers in Fig. 15.

E Visualizing a Training Batch

To better understand our training data mix we randomly sample a batch of 32 image pairs and plot them in Fig. 16.

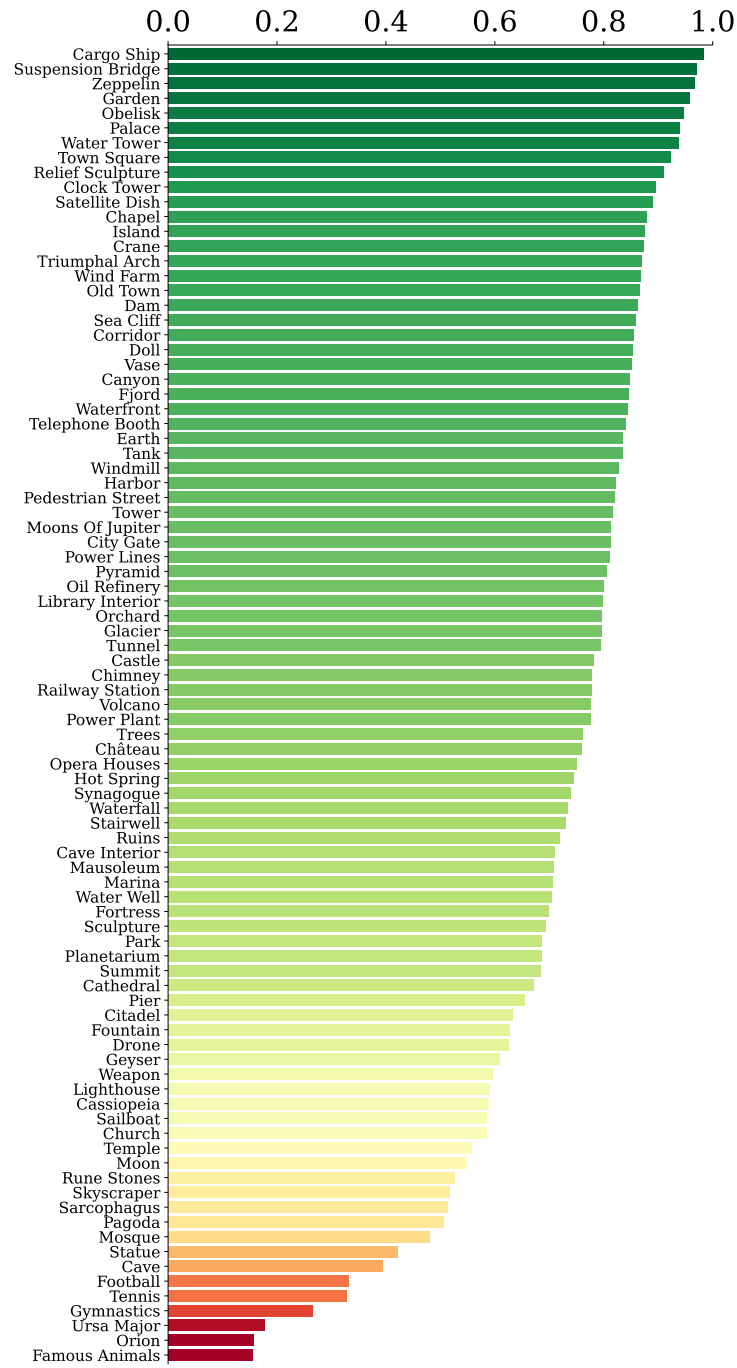


Fig. 14: HardMatch categories from easy to hard. We plot the PCK@10px of LoMa for different categories in the test set.

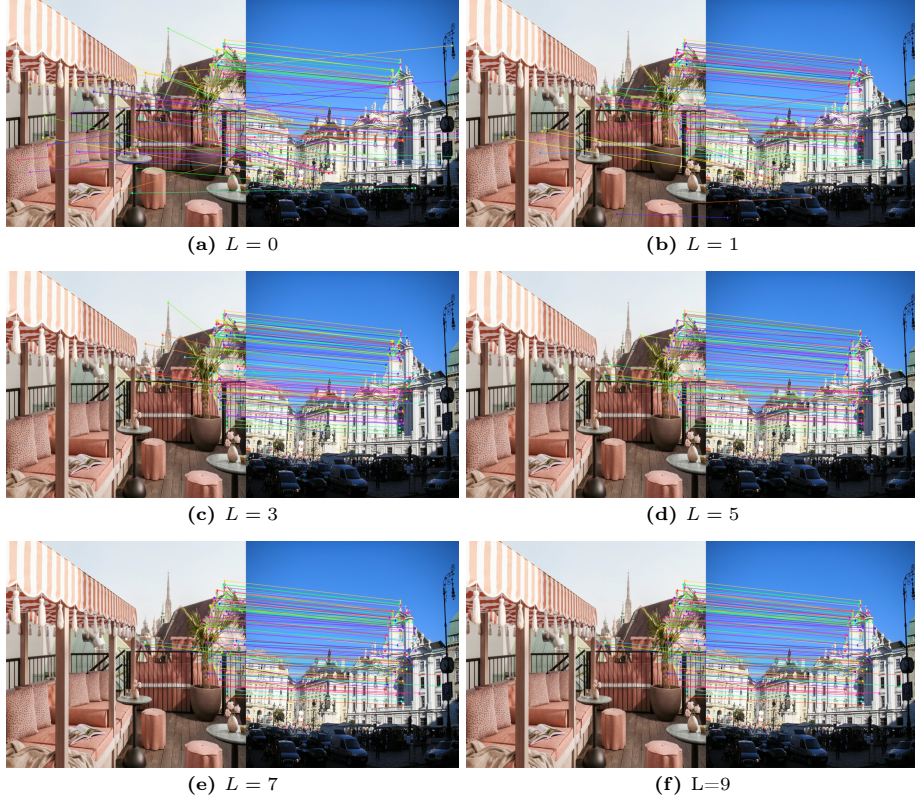


Fig. 15: Refining matches through depth. The descriptor fails to match the pair ($L = 0$) but as the features are passed through the layers of the matcher, the pair gradually becomes matchable.

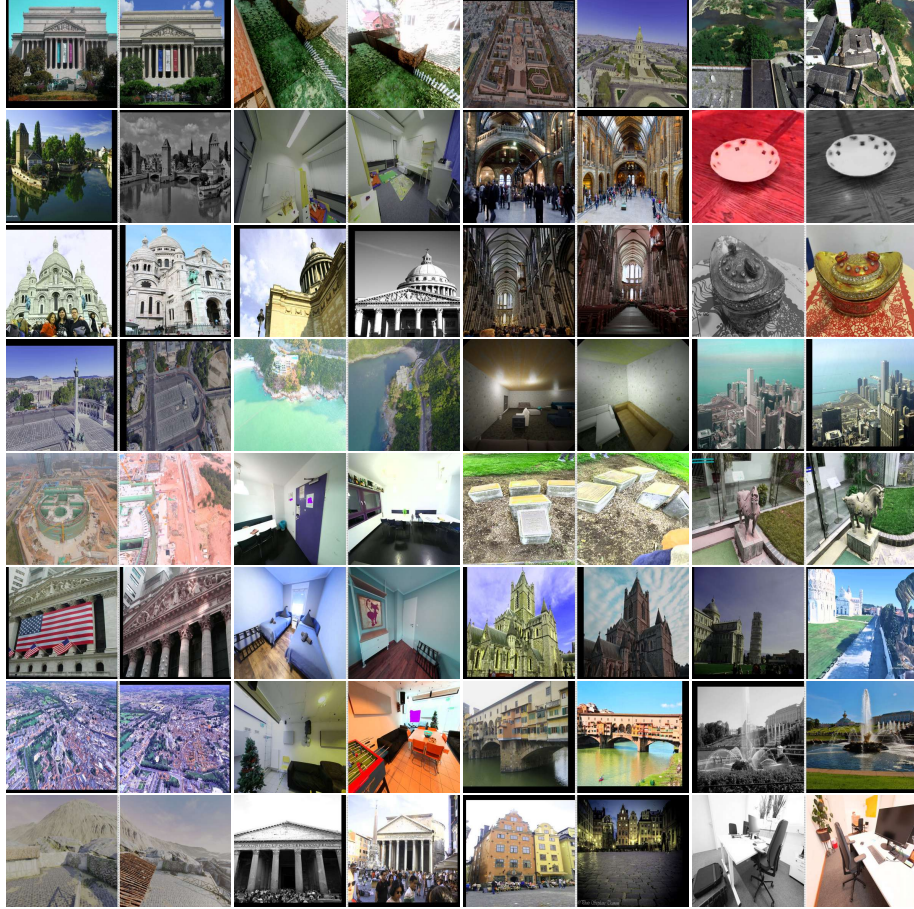


Fig. 16: Visualization of training batch. We visualize a random training batch of 32 image pairs to highlight the diversity in our training data.

References

1. Antequera, M.L., Gargallo, P., Hofinger, M., Buló, S.R., Kuang, Y., Kotschieder, P.: Mapillary planet-scale depth dataset. In: ECCV (2020)
2. Arandjelovic, R., Gronat, P., Torii, A., Pajdla, T., Sivic, J.: Netvlad: Cnn architecture for weakly supervised place recognition. In: CVPR (2016)
3. Arnold, E., Wynn, J., Vicente, S., Garcia-Hernando, G., Monszpart, A., Prisacariu, V., Turmukhambetov, D., Brachmann, E.: Map-free visual relocation: Metric pose relative to a single image. In: ECCV (2022)
4. Avetisyan, A., Xie, C., Howard-Jenkins, H., Yang, T.Y., Aroudj, S., Patra, S., Zhang, F., Frost, D., Holland, L., Orme, C., Engel, J., Miller, E., Newcombe, R., Balntas, V.: Scenescrypt: Reconstructing scenes with an autoregressive structured language model. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds.) ECCV (2025)
5. Balntas, V., Lenc, K., Vedaldi, A., Mikolajczyk, K.: HPatches: A benchmark and evaluation of handcrafted and learned local descriptors. In: CVPR (2017)
6. Barath, D., Matas, J., Noskova, J.: MAGSAC: Marginalizing sample consensus. In: CVPR (2019)
7. Barath, D., Noskova, J., Ivashechkin, M., Matas, J.: Magsac++, a fast, reliable and accurate robust estimator. In: CVPR (2020)
8. Barroso-Laguna, A., Riba, E., Ponsa, D., Mikolajczyk, K.: Key.Net: Keypoint detection by handcrafted and learned cnn filters. In: ICCV (2019)
9. Bökman, G., Edstedt, J., Felsberg, M., Kahl, F.: Steerers: A framework for rotation equivariant keypoint descriptors. In: CVPR (2024)
10. Bökman, G., Nordström, D., Kahl, F.: Flopping for flops: Leveraging equivariance for computational efficiency. In: ICML (2025)
11. Cabon, Y., Murray, N., Humenberger, M.: Virtual kitti 2. arXiv preprint arXiv:2001.10773 (2020)
12. Cai, R., Tung, J., Wang, Q., Averbuch-Elor, H., Hariharan, B., Snavely, N.: Doppelgangsters: Learning to disambiguate images of similar structures. In: ICCV (2023)
13. Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Nießner, M.: Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: CVPR (2017)
14. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: CVPR (2018)
15. Duisterhof, B.P., Zust, L., Weinzaepfel, P., Leroy, V., Cabon, Y., Revaud, J.: Mast3r-sfm: a fully-integrated solution for unconstrained structure-from-motion. In: 3dv (2025)
16. Edstedt, J., Athanasiadis, I., Wadenbäck, M., Felsberg, M.: DKM: Dense kernelized feature matching for geometry estimation. In: CVPR (2023)
17. Edstedt, J., Bökman, G., Zhao, Z.: Dedode v2: Analyzing and improving the dedode keypoint detector. In: CVPRW (2024)
18. Edstedt, J., Bökman, G., Wadenbäck, M., Felsberg, M.: DeDoDe: Detect, Don't Describe – Describe, Don't Detect for Local Feature Matching. In: 3dv (2024)
19. Edstedt, J., Bökman, G., Wadenbäck, M., Felsberg, M.: Dad: Distilled reinforcement learning for diverse keypoint detection. arXiv preprint arXiv:2503.07347 (2025)
20. Edstedt, J., Nordström, D., Zhang, Y., Bökman, G., Astermark, J., Larsson, V., Heyden, A., Kahl, F., Wadenbäck, M., Felsberg, M.: RoMa v2: Harder better faster denser feature matching. In: ECCV (2026)

21. Edstedt, J., Sun, Q., Bökman, G., Wadenbäck, M., Felsberg, M.: RoMa: Robust dense feature matching. In: CVPR (2024)
22. Elflein, S., Zhou, Q., Leal-Taixé, L.: Light3r-sfm: Towards feed-forward structure-from-motion. In: CVPR (2025)
23. Fang, X., Gao, J., Wang, Z., Chen, Z., Ren, X., Lyu, J., Ren, Q., Yang, Z., Yang, X., Yan, Y., Lyu, C.: Dens3r: A foundation model for 3d geometry prediction. In: The Fourteenth International Conference on Learning Representations (2026)
24. Gaidon, A., Wang, Q., Cabon, Y., Vig, E.: Virtual worlds as proxy for multi-object tracking analysis. In: CVPR (2016)
25. Germain, H., Bourmaud, G., Lepetit, V.: S2DNet: learning image features for accurate sparse-to-dense matching. In: ECCV (2020)
26. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge university press (2003)
27. He, X., Sun, J., Wang, Y., Peng, S., Huang, Q., Bao, H., Zhou, X.: Detector-free structure from motion. In: CVPR (2024)
28. Howard, A., Trulls, E., Yi, K.M., Mishkin, D., Dane, S., Jin, Y.: Image matching challenge 2022 (2022), <https://kaggle.com/competitions/image-matching-challenge-2022>
29. Jafarzadeh, A., Antequera, M.L., Gargallo, P., Kuang, Y., Toft, C., Kahl, F., Sattler, T.: Crowddriven: A new challenging dataset for outdoor visual localization. In: ICCV (2021)
30. Jiang, H., Xu, Z., Xie, D., Chen, Z., Jin, H., Luan, F., Shu, Z., Zhang, K., Bi, S., Sun, X., Gu, J., Huang, Q., Pavlakos, G., Tan, H.: Megasynt: Scaling up 3d scene reconstruction with synthesized data. In: CVPR (2025)
31. Keetha, N., Müller, N., Schönberger, J., Porzi, L., Zhang, Y., Fischer, T., Knapitsch, A., Zaus, D., Weber, E., Antunes, N., Luiten, J., Lopez-Antequera, M., Bulò, S.R., Richardt, C., Ramanan, D., Scherer, S., Kotschieder, P.: MapAnything: Universal feed-forward metric 3D reconstruction (2025), arXiv preprint arXiv:2509.13414
32. Lee, J., Yoo, S.: Dense-sfm: Structure from motion with dense consistent matching. In: CVPR (2025)
33. Leroy, V., Cabon, Y., Revaud, J.: Grounding image matching in 3d with mast3r. In: ECCV (2024)
34. Li, Z., Snavely, N.: Megadepth: Learning single-view depth prediction from internet photos. In: CVPR (2018)
35. Lin, H., Chen, S., Liew, J.H., Chen, D.Y., Li, Z., Shi, G., Feng, J., Kang, B.: Depth anything 3: Recovering the visual space from any views. arXiv preprint arXiv:2511.10647 (2025)
36. Lindenberger, P., Sarlin, P.E., Pollefeys, M.: LightGlue: Local Feature Matching at Light Speed. In: ICCV (2023)
37. Liu, C., Yuen, J., Torralba, A.: SIFT flow: Dense correspondence across scenes and its applications. *tpami* **33**(5) (2010)
38. Loiseau, T., Bourmaud, G.: Rubik: A structured benchmark for image matching across geometric challenges. In: CVPR (2025)
39. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: ICLR (2019)
40. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *IJCV* (2004)
41. Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A., Brox, T.: A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In: CVPR (2016)
42. Mishkin, D., Matas, J., Perdoch, M., Lenc, K.: WxBS: Wide baseline stereo generalizations. In: BMVC (2015)

43. Mishkin, D., Radenović, F., Matas, J.: Repeatability is not enough: Learning affine regions via discriminability. In: ECCV (2018)
44. Nordström, D., Edstedt, J., Kahl, F., Bökman, G.: Quick ViTs: Speeding up vision transformers through equivariance. In: ECCV (2026)
45. Nordström, D., Edstedt, J., Bökman, G., Kahl, F.: Who handles orientation? Investigating invariance in feature matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2026)
46. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. In: NeurIPS (2019)
47. Reizenstein, J., Shapovalov, R., Henzler, P., Sbordon, L., Labatut, P., Novotny, D.: Common objects in 3d: Large-scale learning and evaluation of real-life 3d category reconstruction. In: ICCV (2021)
48. Revaud, J., De Souza, C., Humenberger, M., Weinzaepfel, P.: R2d2: Reliable and Repeatable Detector and Descriptor. In: NeurIPS (2019)
49. Roberts, M., Ramapuram, J., Ranjan, A., Kumar, A., Bautista, M.A., Paczan, N., Webb, R., Susskind, J.M.: Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding. In: ICCV (2021)
50. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: An efficient alternative to SIFT or SURF. In: ICCV (2011)
51. Sarlin, P.E., Cadena, C., Siegwart, R., Dymczyk, M.: From coarse to fine: Robust hierarchical localization at large scale. In: CVPR (2019)
52. Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A.: Superglue: Learning feature matching with graph neural networks. In: CVPR (2020)
53. Sattler, T., Maddern, W., Toft, C., Torii, A., Hammarstrand, L., Stenborg, E., Safari, D., Okutomi, M., Pollefeys, M., Sivic, J., Kahl, F., Pajdla, T.: Benchmarking 6dof outdoor visual localization in changing conditions. In: CVPR (2018)
54. Schönberger, J.L., Frahm, J.M.: Structure-from-Motion Revisited. In: CVPR (2016)
55. Su, J., Lu, Y., Pan, S., Murtadha, A., Wen, B., Liu, Y.: Roformer: Enhanced transformer with rotary position embedding (2023), <https://arxiv.org/abs/2104.09864>
56. Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X.: LoFTR: Detector-free local feature matching with transformers. In: CVPR (2021)
57. Taira, H., Okutomi, M., Sattler, T., Cimpoi, M., Pollefeys, M., Sivic, J., Pajdla, T., Torii, A.: Inloc: Indoor visual localization with dense matching and view synthesis. In: CVPR (2018)
58. Tian, Y., Yu, X., Fan, B., Wu, F., Heijnen, H., Balntas, V.: Sosnet: Second order similarity regularization for local descriptor learning. In: CVPR (2019)
59. Tosi, F., Liao, Y., Schmitt, C., Geiger, A.: Smd-nets: Stereo mixture density networks. In: CVPR (2021)
60. Tung, J., Chou, G., Cai, R., Yang, G., Zhang, K., Wetzstein, G., Hariharan, B., Snavely, N.: Megascenes: Scene-level view synthesis at scale. In: ECCV (2024)
61. Tyszkiewicz, M., Fua, P., Trulls, E.: DISK: Learning local features with policy gradient. In: NeurIPS (2020)
62. Verdie, Y., Yi, K., Fua, P., Lepetit, V.: Tilde: A temporally invariant learned detector. In: CVPR (2015)
63. Vuong, K., Ghosh, A., Ramanan, D., Narasimhan, S., Tulsiani, S.: Aerialmegadept: Learning aerial-ground reconstruction and view synthesis. In: CVPR (2025)

64. Wang, B., Chen, C., Cui, Z., Qin, J., Lu, C.X., Yu, Z., Zhao, P., Dong, Z., Zhu, F., Trigoni, N., Markham, A.: P2-net: Joint description and detection of local features for pixel and point matching. In: ICCV (2021)
65. Wang, J., Yuan, Y., Zheng, R., Lin, Y., Gao, J., Chen, L.Z., Bao, Y., Zhang, Y., Zeng, C., Zhou, Y., et al.: Spatialvid: A large-scale video dataset with spatial annotations. arXiv preprint arXiv:2509.09676 (2025)
66. Wang, J., Chen, M., Karaev, N., Vedaldi, A., Rupprecht, C., Novotny, D.: Vggt: Visual geometry grounded transformer. In: CVPR (2025)
67. Wang, J., Chen, M., Zhang, S., Karaev, N., Schönberger, J., Labatut, P., Bojanowski, P., Novotny, D., Vedaldi, A., Rupprecht, C.: Vggt- ω (2026), <https://arxiv.org/abs/2605.15195>
68. Wang, Q.: Understanding and optimizing attention-based sparse matching for diverse local features (2026), <https://arxiv.org/abs/2602.08430>
69. Wang, S., Leroy, V., Cabon, Y., Chidlovskii, B., Revaud, J.: Dust3r: Geometric 3d vision made easy. In: CVPR (2024)
70. Wang, W., Zhu, D., Wang, X., Hu, Y., Qiu, Y., Wang, C., Hu, Y., Kapoor, A., Scherer, S.: Tartanair: A dataset to push the limits of visual slam. In: iros (2020)
71. Wang, Z., Bian, W., Li, X., Tao, Y., Wang, J., Fallon, M., Prisacariu, V.A.: Seeing in the dark: Benchmarking egocentric 3d vision with the oxford day-and-night dataset. In: NeurIPS (2025)
72. Xiangli, Y., Cai, R., Chen, H., Byrne, J., Snavely, N.: Doppelgangers++: Improved visual disambiguation with geometric 3d features. In: CVPR (2025)
73. Yao, Y., Luo, Z., Li, S., Zhang, J., Ren, Y., Zhou, L., Fang, T., Quan, L.: Blend-edmvs: A large-scale dataset for generalized multi-view stereo networks. In: CVPR (2020)
74. Yeshwanth, C., Liu, Y.C., Nießner, M., Dai, A.: Scannet++: A high-fidelity dataset of 3d indoor scenes. In: ICCV (2023)
75. Zhang, Y., Keetha, N., Lyu, C., Jhamb, B., Chen, Y., Qiu, Y., Karhade, J., Jha, S., Hu, Y., Ramanan, D., Scherer, S., Wang, W.: Ufm: A simple path towards unified dense correspondence with flow. In: NeurIPS (2025)
76. Zhao, X., Wu, X., Chen, W., Chen, P.C.Y., Xu, Q., Li, Z.: Aliked: A lighter keypoint and descriptor extraction network via deformable transformation. IEEE Transactions on Instrumentation & Measurement **72** (2023)
77. Zhao, X., Wu, X., Miao, J., Chen, W., Chen, P.C., Li, Z.: Alike: Accurate and lightweight keypoint detection and descriptor extraction. IEEE Transactions on Multimedia (2022)