To Match or Not to Match: Revisiting Image Matching for Reliable Visual Place Recognition - Supplementary Material

Davide Sferrazza[1,2] Gabriele Berton[2] Gabriele Trivigno[2] Carlo Masone[2,3]
[1,2] Politecnico di Torino [3] Focoos AI
[1] {name.surname}@studenti.polito.it [2] {name.surname}@polito.it [3] {name.surname}@focoos.ai

1. Additional Results

In this supplementary, we show:

- Additional qualitative examples of failure cases with high number of inliers in Fig. 1;
- In Fig. 2, the comprehensive set of plots showing the Recall@1 after re-ranking against the average time to process each query over each dataset;
- Uncertainty estimation results (in terms of AUCPR) applied on CliqueMining [2] predictions (the previous SOTA, before MegaLoc [1]), and the re-ranking performance.
- PR curves on the most challenging datasets for CliqueMining;

1.1. Results with CliqueMining

In Tab. 1 we report the uncertainty estimation benchmark starting from CliqueMining predictions, and in Tab. 2 the re-ranking performances of several image matching methods. While CliqueMining in general attains lower performances w.r.t. MegaLoc, these two tables confirm the findings in our main paper. Essentially, higher performances in pure retrieval diminish the usefulness of a matching step. In turn, in scenarios in which the retrieval method struggles, the number of inliers provides a reliable confidence estimate, with AUC scores of 77.1 and 85.7 for the best performing ones on SF-XL Night and Occlusion, respectively. As for MegaLoc, on the more challenging datasets of Baidu, SF-XL Night and Occlusion, the re-ranking step remains beneficial in terms of R@1.

1.2. PR curves for CliqueMining

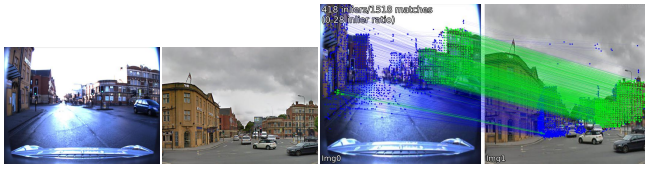
In Fig. 3 we report PR curves for the most challenging datasets of Baidu, SF-XL Night and Occlusion. We show how in these scenarios there is still room for improvement upon CliqueMining predictions, and how image matching methods outperform existing baselines for uncertainty prediction, attaining higher AUC-PR scores.

References

[1] Gabriele Berton and Carlo Masone. Megaloc: One retrieval to place them all. *arXiv preprint arXiv:2502.17237*, 2025. 1

[2] Sergio Izquierdo and Javier Civera. Close, but not there: Boosting geographic distance sensitivity in visual place recognition. In *Computer Vision – ECCV 2024*, pages 240–257, Cham, 2025. Springer Nature Switzerland. 1

(a) SF-XL test V2 – inliers: 1494, distance: 79m



(b) SVOX Sun – inliers: 418, distance: 29m

Figure 1. **Wrong queries with the largest number of inliers for SF-XL test V2 and SVOX Sun.** The displayed images show the query and a confidently retrieved negative image. The dataset name is accompanied by the number of inliers and the distance (in meters) between the two images, as indicated by their labels.

Method	Baidu	MSLS Val	Pitts30k	SF-XL Night	SF-XL Occlusion	SF-XL test V1	SF-XL test V2	SVOX Night	SVOX Sun	Tokyo 24/7	Average
L2-distance	82.7	97.2	98.7	69.0	82.9	98.0	97.9	98.6	99.7	99.9	92.5
PA-Score	85.1	96.7	98.6	63.8	70.7	96.8	97.9	98.7	99.5	99.4	90.7
SUE	88.8	97.3	98.5	73.0	80.4	98.2	**98.6**	99.2	99.6	99.7	93.3
Random	73.1	91.5	92.3	42.4	41.4	83.6	93.1	95.0	97.8	97.2	80.7
R2D2 (NeurIPS '19)	93.7	96.6	98.3	65.0	76.3	98.6	97.9	99.3	99.5	99.7	92.5
D2Net (CVPR '19)	95.1	96.4	98.2	74.1	77.3	99.4	97.9	99.5	99.6	**100.0**	93.7
SuperGlue (CVPR '20)	**95.4**	97.4	98.7	75.0	79.9	99.6	97.8	99.6	99.4	**100.0**	94.3
LoFTR (CVPR '21)	95.1	97.1	98.7	74.0	**85.7**	99.4	98.2	99.6	99.5	**100.0**	**94.7**
Patch2Pix (CVPR '21)	94.0	97.0	98.5	73.0	78.2	99.3	97.7	99.6	99.4	99.9	93.6
Matchformer (ACCV '22)	94.9	97.1	98.8	75.0	84.1	99.5	98.1	**99.7**	99.4	**100.0**	**94.7**
SuperPoint+LightGlue (ICCV '23)	95.0	97.6	98.8	75.3	81.4	99.6	97.9	**99.7**	99.3	**100.0**	94.5
DISK+LightGlue (ICCV '23)	91.8	97.2	98.9	72.2	81.2	99.4	97.7	99.4	99.3	**100.0**	93.7
ALIKED+LightGlue (ICCV '23)	94.3	**97.8**	99.0	73.1	80.2	99.5	98.1	**99.7**	99.6	**100.0**	94.1
RoMa (CVPR '24)	89.7	96.4	96.7	70.4	61.5	97.6	96.1	**99.7**	99.5	**100.0**	90.8
Tiny-RoMa (CVPR '24)	94.8	96.7	98.8	69.6	77.7	98.8	97.4	98.7	99.3	99.8	93.2
Steerers (CVPR '24)	93.9	97.1	**99.1**	72.8	76.8	99.5	98.1	**99.7**	99.2	**100.0**	93.6
Affine Steerers (ECCV '24)	93.4	97.1	98.7	69.7	74.0	99.4	98.1	99.5	99.0	99.9	92.9
DUSt3R (CVPR '24)	89.5	97.4	98.6	69.6	60.8	95.1	96.4	98.7	99.5	99.4	90.5
MASt3R (ECCV '24)	90.4	97.0	**99.1**	72.7	80.1	99.6	97.0	**99.7**	**99.9**	**100.0**	93.5
xFeat (CVPR '24)	92.1	96.9	98.6	70.6	79.5	98.9	97.1	99.2	99.3	99.8	93.2
GIM-DKMv3 (ICLR '24)	84.4	93.0	95.0	67.1	74.5	92.3	94.8	98.3	99.3	99.9	89.9
GIM-LightGlue (ICLR '24)	95.1	97.5	98.8	**77.1**	79.7	**99.7**	97.8	**99.7**	99.5	**100.0**	94.5

Table 1. **The AUPRC of all the baselines and image matching methods**, split according to group type. The shorlist of candidates is obtained with CliqueMining. Distance threshold is fixed at 25 meters. Best overall results on each dataset are in **bold**, best results for each group are underlined.

Method	Baidu R@20 = 96.9		MSLS Val R@20 = 96.6		Pitts30k R@20 = 98.6		SF-XL Night R@20 = 64.8		SF-XL Occlusion R@20 = 65.8		SF-XL test V1 R@20 = 94.0		SF-XL test V2 R@20 = 98.5		SVOX Night R@20 = 99.9		SVOX Sun R@20 = 99.5		Tokyo 24/7 R@20 = 98.4		Average R@20 = 91.3	
	R@1	R@10	R@1	R@10	R@1	R@10	R@1	R@10	R@1	R@10	R@1	R@10	R@1	R@10	R@1	R@10	R@1	R@10	R@1	R@10	R@1	R@10
-	72.9	92.7	**91.6**	**95.9**	**92.6**	97.8	46.1	60.9	44.7	64.5	85.5	92.6	94.5	98.3	**95.5**	99.6	**98.2**	99.4	96.8	97.8	81.8	90.0
R2D2 (NeurIPS '19)	83.2	94.9	82.1	94.7	88.6	97.6	36.9	59.2	46.1	60.5	85.7	92.3	94.1	98.2	80.8	97.3	91.7	98.8	91.1	98.1	78.0	89.2
D2Net (CVPR '19)	83.8	95.1	83.4	95.2	89.4	97.7	44.8	62.4	**53.9**	59.2	88.8	93.4	95.0	**98.5**	86.3	98.2	93.1	98.9	95.6	97.8	81.4	89.6
SuperGlue (CVPR '20)	84.9	95.0	87.7	95.8	89.5	98.0	51.3	**63.5**	48.7	61.8	**91.3**	93.9	94.5	98.3	91.5	98.8	93.8	**99.5**	97.5	97.8	83.1	90.2
LoFTR (CVPR '21)	85.9	95.2	88.1	95.6	90.0	97.6	51.3	62.9	50.0	64.5	91.2	93.5	95.5	**98.5**	93.7	99.4	96.8	99.4	97.5	98.1	84.0	90.5
Patch2Pix (CVPR '21)	83.6	95.2	84.8	95.1	88.8	97.8	41.2	59.9	51.3	61.8	88.8	93.1	94.8	98.3	79.8	99.3	95.3	99.4	96.8	98.1	80.5	89.8
Matchformer (ACCV '22)	85.8	95.1	86.5	95.2	90.7	97.9	50.6	61.6	50.0	63.2	91.0	93.7	**95.7**	**98.5**	94.5	99.1	96.1	**99.5**	**97.8**	**98.4**	83.9	90.2
SuperPoint+LightGlue (ICCV '23)	86.0	95.2	87.3	95.8	90.3	98.0	**53.0**	62.7	47.4	63.2	91.2	**94.0**	93.6	**98.5**	92.2	99.4	95.7	99.3	97.1	98.1	83.4	90.4
DISK+LightGlue (ICCV '23)	84.4	95.4	87.0	95.5	89.3	97.8	50.4	62.4	48.7	64.5	89.1	93.6	92.8	98.3	85.9	97.6	94.1	99.3	96.5	98.1	81.8	90.2
ALIKED+LightGlue (ICCV '23)	86.0	**95.5**	89.2	95.7	89.4	98.0	50.4	**63.5**	51.3	64.5	90.3	93.6	94.5	98.3	88.0	99.0	94.6	99.3	97.1	98.1	83.1	90.5
RoMa (CVPR '24)	81.1	95.0	82.9	95.2	87.4	97.6	42.1	62.0	47.4	63.2	79.4	93.5	93.5	**98.5**	85.4	**99.8**	87.8	**99.5**	93.7	97.8	78.1	90.2
Tiny-RoMa (CVPR '24)	82.4	95.4	82.9	95.2	87.4	97.6	42.1	62.0	47.4	63.2	87.1	93.3	93.5	**98.5**	68.7	97.8	82.7	99.2	93.7	97.8	76.8	90.0
Steerers (CVPR '24)	85.7	95.0	81.4	94.4	89.7	97.7	44.8	61.8	50.0	61.8	90.3	93.7	95.0	**98.5**	88.5	99.3	95.3	99.3	96.5	**98.4**	81.7	90.0
Affine Steerers (ECCV '24)	84.4	95.1	84.2	95.3	89.1	97.7	47.4	62.4	48.7	59.2	89.5	93.7	94.5	98.2	88.5	99.3	94.3	99.1	96.8	98.1	81.7	89.8
DUSt3R (CVPR '24)	77.2	94.7	75.3	89.5	83.0	95.6	35.0	53.9	36.8	52.6	77.1	87.0	76.9	97.7	67.4	79.1	77.9	86.2	88.3	96.8	69.5	83.3
MASt3R (ECCV '24)	83.4	95.3	80.5	95.7	90.9	**98.1**	48.1	62.2	51.3	**65.8**	90.3	93.8	93.5	98.3	95.3	99.6	96.3	99.4	96.8	**98.4**	82.6	**90.7**
xFeat (CVPR '24)	81.4	94.1	85.8	95.0	89.4	97.7	43.6	61.2	43.4	60.5	86.9	93.0	93.5	**98.5**	83.1	98.2	89.0	98.6	93.0	97.8	78.9	89.5
GIM-DKMv3 (ICLR '24)	53.0	95.2	32.1	92.3	66.2	97.7	38.2	62.4	36.8	63.2	53.8	93.5	75.3	98.0	55.2	**99.8**	51.8	99.3	77.8	98.1	54.0	90.0
GIM-LightGlue (ICLR '24)	**86.5**	95.4	88.6	95.7	91.1	97.8	51.5	61.4	52.6	64.5	91.1	93.8	94.3	98.3	93.1	97.8	96.8	98.9	96.5	98.1	**84.2**	90.2

Table 2. **Recalls before and after applying re-ranking.** Recalls are computed by setting the distance threshold to 25 meters. The shortlist of candidates to be re-ranked is obtained with CliqueMining, and the results with such shortlist are shown in the first row. Re-ranking has been applied to the first 20 candidates (*i.e.* $K = 20$). Next to each dataset's name, we show the R@20, which in practice sets the upper bound of the maximum recalls achievable after re-ranking. Best results are in **bold**, second best are underlined.

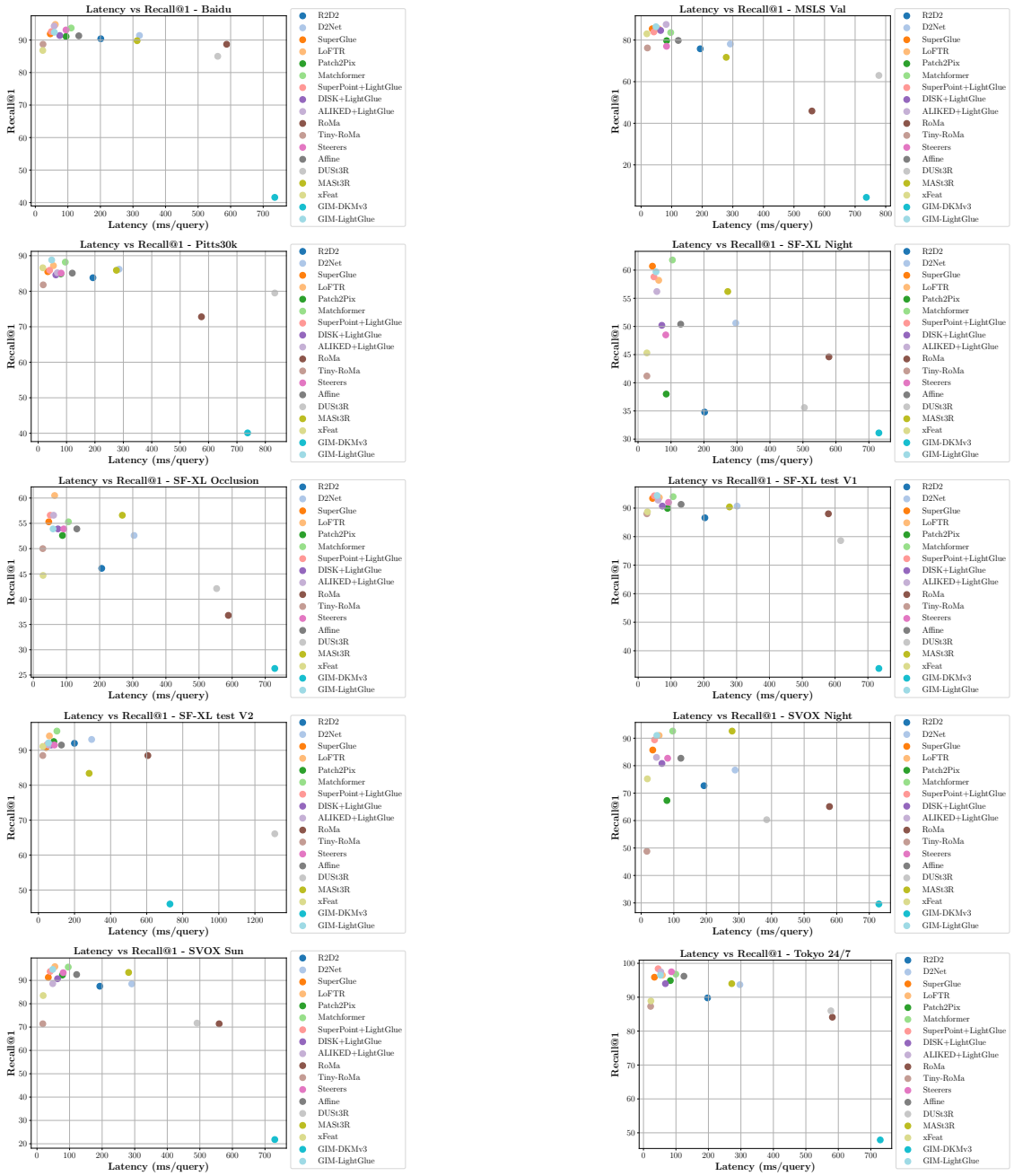


Figure 2. **Plot displaying the Recall@1 after re-ranking and mean latency for different methods.** For each dataset, we compute the Recall@1 after re-ranking and the mean latency, which is the average time to process each query. The shortlist of candidates for the Recall@1 is obtained with MegaLoc and distance threshold fixed at 25 meters.

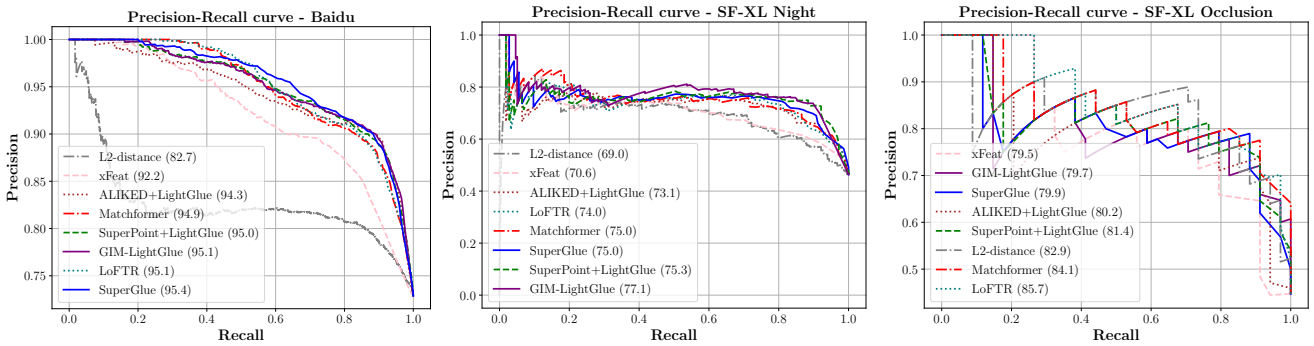


Figure 3. **Precision-Recall curves**, computed for the top-7 image matching methods on Baidu, SF-XL Night, and SF-XL Occlusion, together with the L2 distance in embedding space. Distance threshold is fixed at 25 meters.