Improving MegaLoc retrieval in combination with Superpoint and Superglue for CroCoDL localization challenge' 25

Huy-Hoang Bui*†
Woven by Toyota
Tokyo, Japan
hoang.bh15031999@gmail.com

Danh-Khang Cao*†
Viettel High Tech
Hanoi, Vietnam
danhkhang233@gmail.com

Gabriele Berton*†
Amazon
Palo Alto, USA
gabriele.berton@email.com

Abstract

In this report, we present our method for achieving precise localization results in the CroCoDL Localization 2025 Challenge. We propose a localization pipeline built on top of Hierarchical Localization, combining several models as follows: ***a)*** *global feature extraction is performed using MegaLoc with a DINOv2-L backbone, combined with Mast3R for re-ranking;* ***b)*** *SuperPoint and RDD are used for feature extraction;* ***c)*** *SuperGlue and LightGlue are used for feature matching, depending on the scene.*

1. Approach

We first established a baseline using the default configuration and model provided in the `croco_benchmark` code base. We used a combination of Megaloc[2], SuperPoint[4], and SuperGlue[7] with their default settings. The number of retrieved images was set to 10 for both triangulation and localization. The results of this pipeline are shown in Table 1, with an average localization accuracy of 87.47%. Based on this baseline, we applied the following techniques to improve localization performance.

We changed the number of retrieved images for localization. As the number increased, localization accuracy improved, but this also increased the computation time for image matching and especially for RANSAC PnP. We chose to use 30 retrieved images for localization, as it provided accurate results while keeping the runtime short. From our observation, testing with 40 or 50 retrieved images took much longer to compute poses and gave only minor improvements.

At this stage, we identified difficult cases mostly concentrated in `SUCCULENT` are `spot_map/ios_query`, `spot_map/hl_query`, `hl_map/ios_query`, `ios_map/ios_query`, `hl_map/spot_query`.

Table 1. Baseline pipeline performance on HYDRO and SUCCULENT.

(a) HYDRO

		map		
		iOS	HL	Spot
query	iOS	89.43%	96.31%	74.20%
	HL	92.55%	98.69%	92.55%
	Spot	83.99%	96.46%	98.88%

(b) SUCCULENT

		map		
		iOS	HL	Spot
query	iOS	80.95%	77.12%	65.74%
	HL	90.49%	95.44%	70.71%
	Spot	85.79%	85.05%	100.00%

The default MegaLoc uses DINOv2 [6] ViT/B-14 as its backbone. We replaced it with a larger backbone, ViT/L-14. This change brought significant improvements, especially in difficult scenes such as `SUCCULENT`, `spot_map/ios_query`, `hl_map/ios_query`, and `hl_map/spot_query`. In addition, after obtaining the list of retrieved images using the new backbone, we re-ranked them by running image matching between the queries and the retrieved images using Mast3R [5]. The retrieved images were then sorted by the number of inliers computed from 2D-2D matches using RANSAC.

For feature detection and matching, we tested several detector–matcher pairs, including ALIKED[8]/LightGlue and RDD[3]/LightGlue. On the `SUCCULENT` `ios_map/ios_query` scene, RDD with LightGlue achieved higher performance compared to the default SuperPoint/SuperGlue setup. The RDD configuration was adopted from the 4th place solution of the Image Matching

*All authors contributed equally to this submission.
†This work was conducted in author personal time and not related to employment at author's organization.

Challenge 2025 [1].

2. Results and discussion

Table 2. Combined best results on HYDRO and SUCCULENT.

(a) HYDRO

		map		
		iOS	HL	Spot
query	iOS	92.87%	99.26%	85.26%
	HL	93.36%	98.98%	96.13%
	Spot	87.57%	96.65%	98.88%

(b) SUCCULENT

		map		
		iOS	HL	Spot
query	iOS	85.58%	85.19%	75.40%
	HL	92.96%	94.69%	76.81%
	Spot	90.72%	88.51%	100.00%

Running the entire dataset is time-consuming, and we are limited in compute resources, so we conducted experiments on selected scenes and combined their results, as shown in Table 2.

From the baseline results in Table 1, the average accuracy was 87.47%. By increasing the number of retrieved images for localization to 30, the pipeline reached 89.20% accuracy. Switching to DINOv2 ViT/L-14 and RDD/LightGlue for `SUCCULENT ios_map/ios_query` further improved the overall accuracy to 90.59%. Finally, applying Mast3R re-ranking increased the accuracy to 91.05%. Details of the best combined results from all experiments are shown in Table 2.

3. Conclusion

For this challenge, we applied multiple techniques for image retrieval, feature extraction, and image matching that showed promising localization results. The entire pipeline can run on a consumer GPU and CPU, making it very efficient. However, we did not have enough time to explore more approaches for cases with high domain shift, such as `SUCCULENT spot_map/ios_query` and `spot_map/hl_query`. Some initial experiments showed good results using line matching, reconstruction, and learning-based techniques, which we plan to explore further in the future.

References

[1] Fabio Bellavia, Jiri Matas, Dmytro Mishkin, Luca Morelli, Fabio Remondino, Amy Tabb, Eduard Trulls, Kwang Moo Yi, Sohier Dane, Addison Howard, and María Cruz. Image matching challenge 2025. https://www.kaggle.com/competitions/image-matching-challenge-2025, 2025. Kaggle. Accessed: 2025-10-09. 2

[2] Gabriele Berton and Carlo Masone. Megaloc: One retrieval to place them all. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2861–2867, 2025. 1

[3] Gonglin Chen, Tianwen Fu, Haiwei Chen, Wenbin Teng, Hanyuan Xiao, and Yajie Zhao. Rdd: Robust feature detector and descriptor using deformable transformer. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6394–6403, 2025. 1

[4] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 224–236, 2018. 1

[5] Vincent Leroy, Yohann Cabon, and Jérôme Revaud. Grounding image matching in 3d with mast3r. In *European Conference on Computer Vision*, pages 71–91. Springer, 2024. 1

[6] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 1

[7] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 4938–4947, 2020. 1

[8] 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*, 72:1–16, 2023. 1