Robust Scene Coordinate Regression via Geometrically-Consistent Global Descriptors

Supplementary Material

Son Tung Nguyen Alejandro Fontan Michael Milford Tobias Fischer

Queensland University of Technology
Brisbane, Australia

sontung.nguyen@hdr.qut.edu.au

1. Margin τ in the mGCL function

We analyze the impact of the margin parameter τ in the mGCL loss (Eq. (2) in the main paper). As shown in Figure 1, localization accuracy increases with larger margins, peaking at $\tau = 0.9$ with an improvement of around $2-3\%$. In our experiments, we use $\tau = 0.5$ as the default.

2. PCA dimension for compressing local descriptors

We analyze the impact of the PCA dimension used to compress local descriptors. As shown in Figure 2, aggressive compression reduces localization accuracy, whereas performance remains stable for PCA dimensions between 64 and 128. We use 128 as the default to enable fair comparison with R-Score [6].

3. The Aachen Day/Night v1.1 benchmark

We further evaluate our method on the Aachen Day/Night v1.1 dataset [9], which extends the original benchmark with broader inner-city coverage, increasing visual ambiguity. As expected, accuracy decreases for both R-Score [6] and our method; however, our approach still outperforms R-Score, with an average gain of 2.3% (Table 1).

4. Runtime

Our method has comparable runtime to R-Score [6] because both rely on the same coordinate encoding. The main difference lies in the global descriptor: R-Score uses the heavier NetVLAD [1], while we adopt the more efficient SALAD [5]. This yields higher overall efficiency for our system (see Table 2).

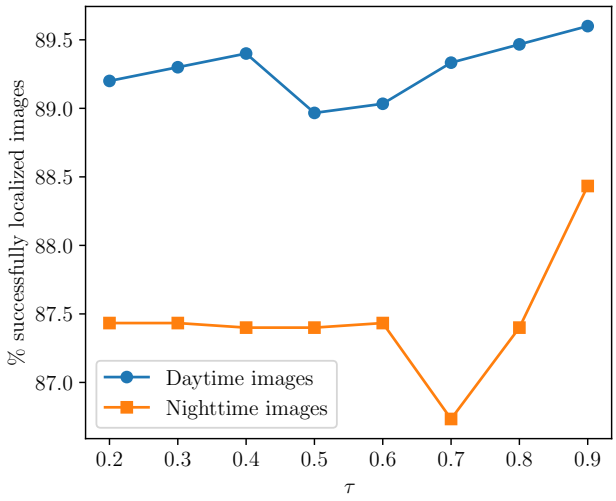

Figure 1. **Margin τ.** We analyze the impact of the margin parameter τ on localization accuracy. Larger margins typically produce more discriminative descriptors and improved accuracy, with the best performance achieved at $\tau = 0.9$, resulting in a gain of approximately $2-3\%$. For consistency across experiments, we set $\tau = 0.5$ as the default.

5. Local descriptors

We conduct an ablation to assess the impact of different local descriptors, evaluating SuperPoint [2], D2-Net [3], and DeDoDe [4]. The results in Table 3 show that DeDoDe performs substantially better, exceeding the other methods by at least 40% under the 0.25m / 2° criterion on the Aachen nighttime images.

References

[1] Relja Arandjelovic, Petr Gronát, Akihiko Torii, Tomás Pajdla, and Josef Sivic. Netvlad: CNN architecture for weakly super-

	Memory requirement ↓	Aachen day			Aachen night			Average ↑ (%)
		0.25m/2°	0.5m/5°	5m/10°	0.25m/2°	0.5m/5°	5m/10°	
hLoc (SP+SG) [7, 8]	> 7.82 GB	89.8	96.1	99.4	77.0	90.6	100.0	92.1
R-Score [6]	47 MB	76.8	90.0	97.6	53.4	82.2	97.9	83.0
Ours	53 MB	82.8	93.1	98.8	53.4	84.8	99.0	85.3

Table 1. **Aachen day/night v1.1 Dataset [9].** We report the percentage of query images successfully localized under varying error thresholds. Under the 0.25m / 2° threshold, our system outperforms R-Score [6] by 6.0% under day conditions, respectively.

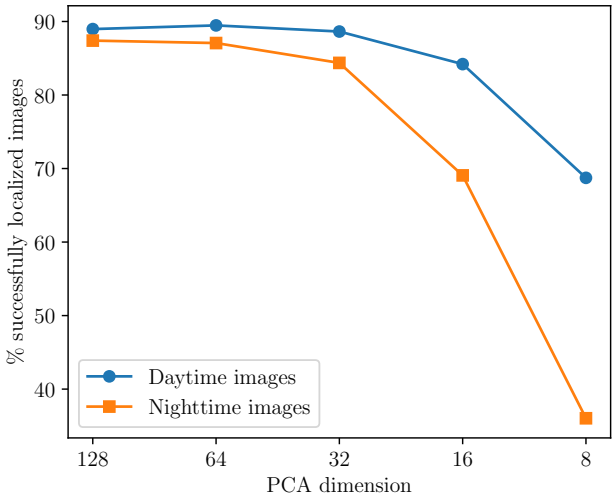

Figure 2. **PCA dimension.** We analyze the impact of the PCA dimension on localization accuracy. Localization accuracy decreases as compression becomes more aggressive. Nevertheless, the accuracy remains fairly consistent when the PCA dimension is between 64 and 128. In our experiments, we adopt a PCA dimension of 128 as the default setting to ensure a fair comparison with R-Score [6].

	Aachen day/night	Dept. B1	Dept. 1F	Dept. 4F
R-Score	4.35	2.29	2.16	2.21
Ours	5.46	2.42	2.89	2.28

Table 2. **Frames per second.** We evaluate the frame rate of our method and R-Score [6]. While both approaches use the same coordinate encoding, our method employs the more efficient SALAD [5] for global descriptor extraction instead of NetVLAD [1]. This leads to improved overall runtime efficiency for our approach.

	Aachen Day			Aachen Night		
Ours (D2-Net [3])	36.5	55.9	82.4	8.2	17.3	55.1
Ours (SuperPoint [2])	73.8	84.3	95.4	30.6	61.2	85.7
Ours (ALIKED [10])	64.4	80.9	94.3	45.9	69.4	98.0
Ours (DeDoDe [4])	**80.3**	**90.3**	**97.1**	**72.4**	**90.8**	**99.0**

Table 3. **Ablation study of local descriptors.** We include the ablation studies of different local descriptors for our system. DeDoDe [4] outperforms other methods [2,3] by a large margin, hence being the default choice.

vised place recognition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5297–5307, 2016. 1, 2

[2] 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, 2

[3] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8092–8101, 2019. 1, 2

[4] Johan Edstedt, Georg Bökman, Mårten Wadenbäck, and Michael Felsberg. Dedode: Detect, don't describe—describe, don't detect for local feature matching. In *Int. Conf. 3D Vision*, pages 148–157, 2024. 1, 2

[5] Sergio Izquierdo and Javier Civera. Optimal transport aggregation for visual place recognition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 17658–17668, 2024. 1, 2

[6] Xudong Jiang, Fangjinhua Wang, Silvano Galliani, Christoph Vogel, and Marc Pollefeys. R-SCoRe: Revisiting scene coordinate regression for robust large-scale visual localization. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 11536–11546, 2025. 1, 2

[7] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 12716–12725, 2019. 2

[8] 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. 2

[9] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8601–8610, 2018. 1, 2

[10] Xiaoming Zhao, Xingming Wu, Weihai Chen, Peter C. Y. Chen, Qingsong Xu, and Zhengguo Li. Aliked: A lighter keypoint and descriptor extraction network via deformable transformation. *IEEE Transactions on Instrumentation & Measurement*, 72:1–16, 2023. 2