Robust scene coordinate regression for large-scale visual localization

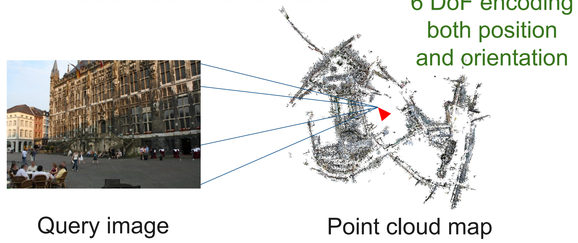
Son Tung Nguyen, Alejandro Fontan, Michael Milford, Tobias Fischer
QUT Centre for Robotics, Queensland University of Technology

Centre for Robotics

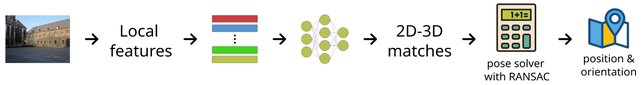
WACV 2026 TUCSON, AZ 3/6 – 3/10

Abstract

The visual localization task



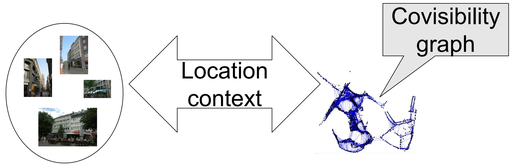
Scene coordinate regression (SCR) models



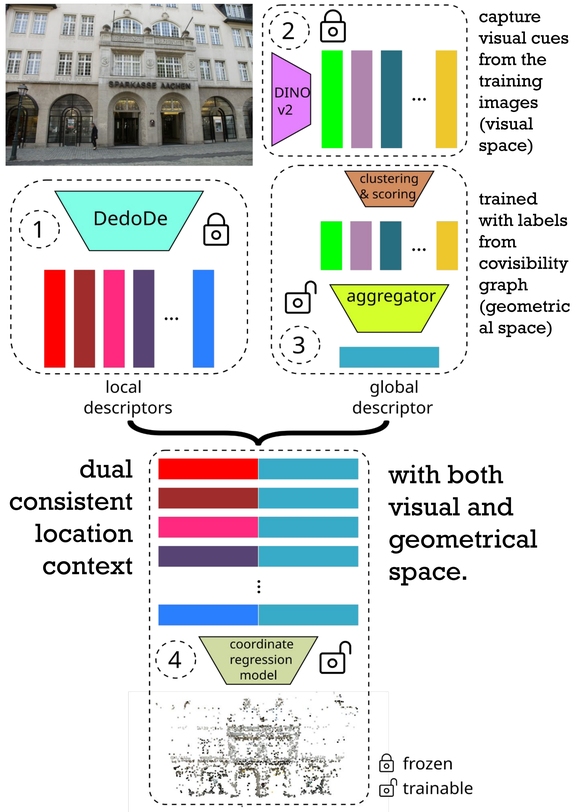
Literature gap

Location context for SCR models:

- Pretrained global descriptors such as GLACE [1] are not smooth enough for larger maps.
- Covisibility graph embeddings as in R-Score [2] skip visual cues and lead to noisy graph connections
- This work offer slocation context that is **consistent with both image space and covisibility graph**.

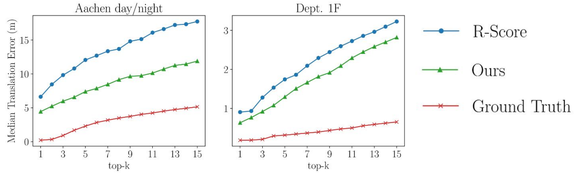


System overview



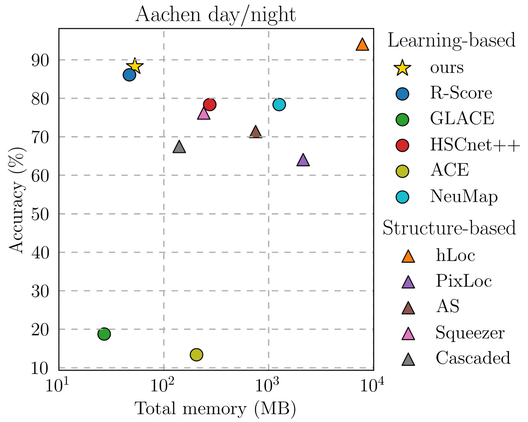
Results

Location context quality



- Similar places (by translation difference) should have similar location context in the global descriptor space.
- Better retrieval performance compared to raw covisibility graph embeddings (lower translation differences for top-k retrievals).

Quantitative results



- 2.5% higher accuracy on average on Aachen day/night dataset, with only 7 MB additional storage.

Ablation studies

	Average accuracy
NetVLAD + GCL [5]	81.2%
FocusTune [3] + NetVLAD + GCL [5]	86.3%
FocusTune [3] + SALAD [4] + GCL [5]	87.3%
FocusTune [3] + SALAD [4] + mGCL	**88.3%**

- Buffer sampling matters: FocusTune [3] improves the accuracy by 5.5%.
- Accuracy improvement by using better global descriptors at query time: SALAD [4] improves the accuracy by 1%.
- Using our proposed mGCL loss with smoother distance function is better than GCL [5].

SCAN QR FOR THE CODE!

Contact
Son Tung Nguyen
Queensland University of Technology
Email: tung.nguyenson1404@gmail.com

Acknowledgements
This research was partially supported by the QUT Centre for Robotics, an ARC Laureate Fellowship FL210100156 to Michael Milford, and an ARC DECRA Fellowship DE240100149 to Tobias Fischer.

References

[1] GLACE: Global Local Accelerated Coordinate Encoding, Wang et al., CVPR 2024.
[2] R-SCoRe: Revisiting Scene Coordinate Regression for Robust Large-Scale Visual Localization, Jiang et al., CVPR 2025.
[3] FocusTune: Tuning Visual Localization through Focus-Guided Sampling, Nguyen et al., WACV 2024.
[4] Optimal Transport Aggregation for Visual Place Recognition, Izquierdo and Civera, CVPR 2024.
[5] Data-efficient Large Scale Place Recognition with Graded Similarity Supervision, Leyva-Vallina et al., CVPR 2023.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com