# Supplementary Material
# CASSPR: Cross Attention Single Scan Place Recognition

Yan Xia*[1,2,3†] Mariia Gladkova*[1,5] Rui Wang[4] Qianyun Li[1] Uwe Stilla[1] João F. Henriques[3] Daniel Cremers[1,2,3,5]
[1]Technical University of Munich [2]Munich Center for Machine Learning (MCML)
[3] Visual Geometry Group, University of Oxford [4] Microsoft Zurich [5] Munich Data Science Institute
{yan.xia, mariia.gladkova, stilla, cremers}@tum.de, qianyunli0@outlook.com, joao@robots.ox.ac.uk

## A. Overview

In this supplementary material, we provide more experiments on the USyd Campus [2] and Oxford RobotCar [1] datasets to demonstrate the effectiveness of our CASSPR and show more insights we gathered during the development.

## B. Number of lightweight self-attention units

In this section, we explore the network performance with different numbers of lightweight self-attention (LSA) units on the Oxford RobotCar [1] and in-house datasets. Specifically, we insert LSA units one by one after each convolutional layer in CASSPR. The network is denoted as CASSPR_LSA. '0' means that we do not add any LSA units.

Table 1 shows results of average recall at top 1% and top 1 with different numbers of LSA units for the CASSPR_LSA architecture. As seen from the table, CASSPR_LSA achieves the best performance with 6 LSA units. This implies the network has the best global awareness when all attention units are used.

## C. The maximum range of 3D LiDAR scans

In this section, we test the performance of the proposed CASSPR with different maximum ranges of points from LiDAR. The experiments are conducted on the USyd dataset [2]. We set the maximum range of points from 20 m to 100 m at 20 m intervals since the Velodyne VLP-16 sensor utilized in the USyd dataset has a range of about 100 m. Fig. 1 presents the results, the best performance is obtained when the maximum measurement range is set to at least 60m.

We can get two conclusions:(1) Compared to the performance from 40 m to 100 m, our method achieves fairly good average recall at top 1 ranging from 85.0% to 87.8%, demonstrating our CASSPR is robust to the different maximum ranges of points. (2) When the maximum range is 20 m, the performance at top 1 drops to 77.7% due to the limited information provided by the single scan within a small range.

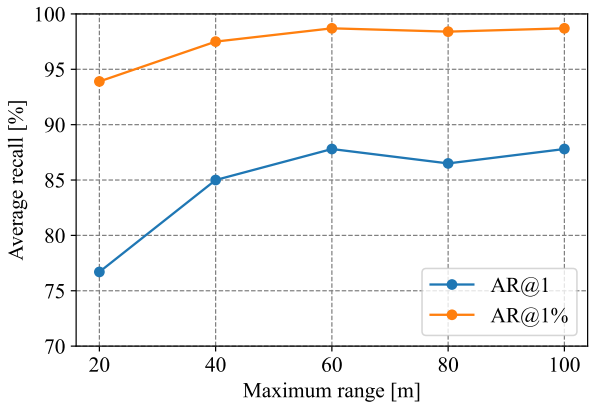


Figure 1. Average recall of CASSPR tested on the USyd with a different maximum distance of points from the scanner position. Maximum ranges less than 40 m show a drop in performance.

## References

[1] Will Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 1 year, 1000 km: The oxford robotcar dataset. *The International Journal of Robotics Research*, 36(1):3–15, 2017. 1

[2] Wei Zhou, Julie Stephany Berrio, Charika De Alvis, Mao Shan, Stewart Worrall, James Ward, and Eduardo Nebot. Developing and testing robust autonomy: The university of sydney campus data set. *IEEE Intelligent Transportation Systems Magazine*, 12(4):23–40, 2020. 1

†Corresponding author. * Equal contribution.

Table 1. Average recall (%) at top 1% (@1%) and top 1 (@1) for CASSPR_LSA with different numbers of LSA units trained only on the Oxford RobotCar.

| Number of LSA | Oxford RobotCar | | U.S. | | R.A. | | B.D. | |
|---|---|---|---|---|---|---|---|---|
| | AR @1 | AR @1% | AR @1 | AR @1% | AR @1 | AR @1% | AR @1 | AR @1% |
| 0 | 93.0 | 97.9 | 86.7 | 95.0 | 80.4 | 91.2 | 81.5 | 88.5 |
| 1 | 93.5 | 98.0 | 86.9 | 95.0 | 80.0 | 88.8 | 80.9 | 87.3 |
| 2 | 93.5 | 98.0 | 87.2 | 96.0 | 80.2 | 89.3 | 80.8 | 87.4 |
| 3 | 92.9 | 97.6 | 90.1 | 96.5 | 81.0 | 91.1 | 82.6 | 89.2 |
| 4 | 92.9 | 97.6 | 87.2 | 95.6 | 82.7 | 92.5 | 82.5 | 89.1 |
| 5 | **94.7** | **98.4** | 88.0 | 94.4 | 86.2 | 93.4 | 82.7 | 89.0 |
| 6 | **94.7** | **98.4** | **91.4** | **97.1** | **86.8** | **93.5** | **86.3** | **91.0** |