CS-Wild-Places: A Novel Benchmark for Cross-Source Lidar Place Recognition in Forest Environments

Ethan Griffiths[1,2] Maryam Haghighat[1] Simon Denman[1] Clinton Fookes[1]
Milad Ramezani[2]

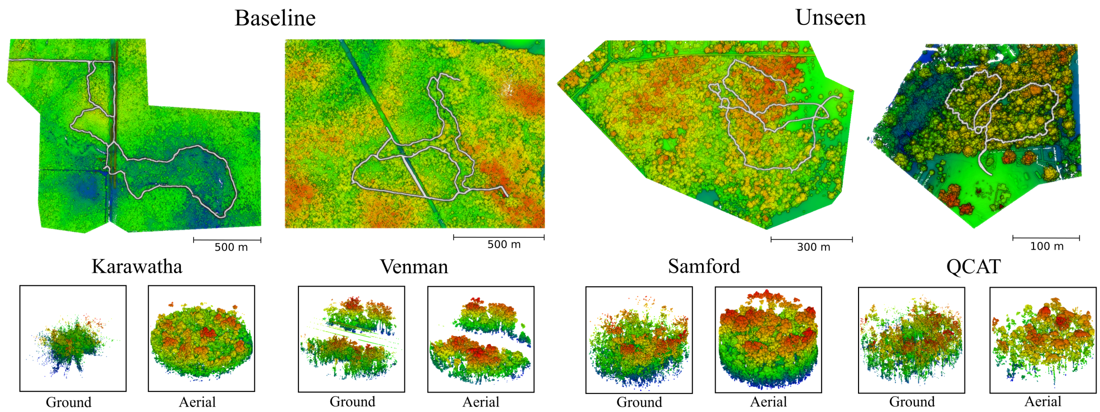

Figure 1: Overview of the CS-Wild-Places dataset. (Top row) bird's eye view of aerial global maps from each forest with ground trajectory overlaid. (Bottom row) sample ground and aerial submap from each forest.

This document outlines the structure of the CS-Wild-Places dataset, as well as instructions for downloading and using the dataset. The dataset, code, and associated HOTFormerLoc [1] paper can be found at https://csiro-robotics.github.io/HOTFormerLoc.

CS-Wild-Places contains six lidar sequences (four aerial and two ground) captured in four forests across Brisbane: Karawatha Forest Park, Venman Bushland National Park, QCAT Forest, and Samford Ecological Research Facility, as shown in Fig. 1. These sequences supplement the eight ground lidar sequences proposed in the Wild-Places [2] dataset, to produce a comprehensive set of geo-registered ground and aerial lidar submaps suitable for large-scale cross-source place recognition research.

[1]Ethan Griffiths, Maryam Haghighat, Simon Denman, and Clinton Fookes are with the Signal Processing, Artificial Intelligence and Vision Technologies (SAIVT) lab at the Queensland University of Technology (QUT), Brisbane, Australia. Emails: {maryam.haghighat, s.denman, c.fookes}@qut.edu.au

[2]Ethan Griffiths and Milad Ramezani are with CSIRO Robotics, Data61, CSIRO, Brisbane, Australia. Emails: {ethan.griffiths, milad.ramezani}@data61.csiro.au

1 Dataset Structure

```
CS-Wild-Places/
├── postproc_voxel_0.80m_rmground_normalised/
│   ├── Karawatha/
│   │   ├── 2021_06_21_K-02_ground_sample0.5s/
│   │   ├── 2021_06_22_K-01_ground_sample0.5s/
│   │   ├── 2021_12_13_K-03_ground_sample0.5s/
│   │   ├── 2022_08_11_K-04_ground_sample0.5s/
│   │   └── 2024_07_10_aerial_sample10m/
│   │       ├── clouds/
│   │       │   ├── <timestamp00>.pcd
│   │       │   ├── <timestamp01>.pcd
│   │       │   ├──         ⋮
│   │       │   └── <timestampXX>.pcd
│   │       └── poses.csv
│   ├── QCAT/
│   │   ├── 2023_09_07_ground_sample2s/
│   │   └── 2023_09_08_aerial_sample10m/
│   ├── Samford/
│   │   ├── 2023_09_17_ground_sample2s/
│   │   └── 2023_09_18_aerial_sample10m/
│   └── Venman/
│       ├── 2021_06_11_V-01_ground_sample0.5s/
│       ├── 2021_06_11_V-02_ground_sample0.5s/
│       ├── 2021_12_16_V-03_ground_sample0.5s/
│       ├── 2022_08_12_V-04_ground_sample0.5s/
│       └── 2024_07_09_aerial_sample10m/
├── postproc_voxel_0.80m_rmground/
│   ├── Karawatha/
│   ├── QCAT/
│   ├── Samford/
│   └── Venman/
└── raw_500k_max/
    ├── Karawatha/
    ├── QCAT/
    ├── Samford/
    └── Venman/
```

Figure 2: File structure of CS-Wild-Places. Sequences highlighted in red are from the Wild-Places [2] dataset, which must be downloaded and processed separately.

Fig. 2 outlines the file structure of CS-Wild-Places. We offer submaps in three formats: raw submaps limited to 500k points (`raw_500k_max/`), post-processed with ground points removed and downsampled with voxel size 0.80 m (`postproc_voxel_0.80m_rmground/`), and the same with point coordinates normalised to $[-1, 1]$ (`postproc_voxel_0.80m_rmground_normalised/`). Note that all submaps are cropped to a horizontal radius of 30 m. In our experiments, we use the `postproc_voxel_0.80m_rmground_normalised/` submaps. We also provide the original script used for post-processing CS-Wild-Places submaps here, with support for multiple downsampling methods, so users can post-process CS-Wild-Places to suit their needs. See the HOTFormerLoc paper [1] for further details on submap generation.

Individual sequences are stored in the format `<date>_<name>_<type>_sample<XX>/` (note that `<name>` is only specified for Wild-Places sequences for clarity). This indicates the date each sequence was captured, the type of sequence (ground or aerial), and the sampling frequency (seconds between samples for ground, and metres between grid points for aerial). Also note that sequences from Wild-Places [2] (e.g. `2021_06_21_K-02_ground_sample0.5s/`) do not come packaged with our dataset, and must be downloaded and processed as detailed in Section 2.

1.1 Submaps and Pose Files

Within each sequence, submaps are stored in the `clouds/` directory in `.pcd` format with points in the sensor coordinate frame, which can be loaded using the PCL, Open3D, or Pyntcloud libraries. 6-DoF poses for each submap are stored in the `poses.csv` file in quaternion format with UTM Easting and Northing coordinates (UTM zone 56S). Individual submaps are named according to their corresponding timestamp, which can be used to find the associated 6-DoF pose from the pose file. Note that aerial submaps have pseudo-timestamps due to the grid-based sampling approach used, and thus are stored with filenames based on their index position within the grid spanning the aerial map (e.g. `10000Y.0000X`, where `Y` and `X` are the grid indices in Northing and Easting directions, respectively).

2 Downloading and Preparing CS-Wild-Places

The CS-Wild-Places dataset can be downloaded from the CSIRO Data Access Portal (DAP). Below is a step-by-step guide for downloading and processing CS-Wild-Places.

1. Navigate to the "Files" tab in the DAP and select the folders you wish to download, as shown in Fig. 3.

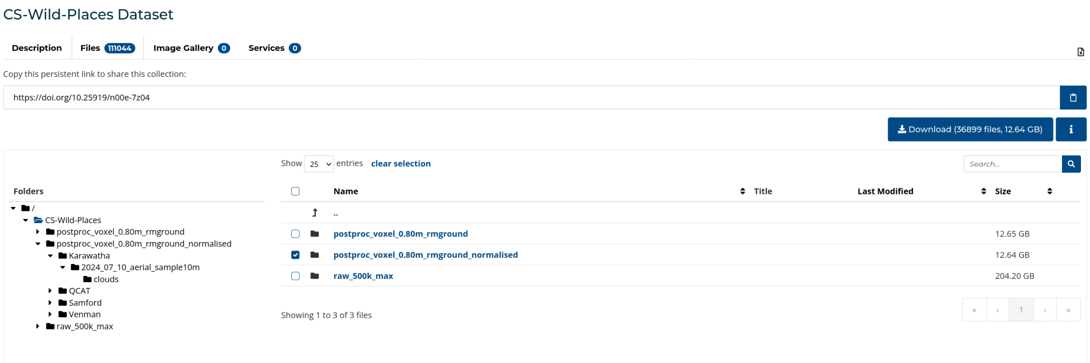

Figure 3: Preview of the CSIRO Data Access Portal (DAP) page.

2. Select the "Download" button in the top-right, and select the "Download files via S3 Client" option in the drop-down box. A window similar to Fig. 4 will appear, providing instructions for downloading via an S3 compatible client. If you have a suitable S3 client, use these instructions, otherwise navigate to the "rClone" or "AWS CLI" tabs if you prefer

to download via a single command entered on the command-line. Download CS-Wild-Places to a suitable directory of your choosing.

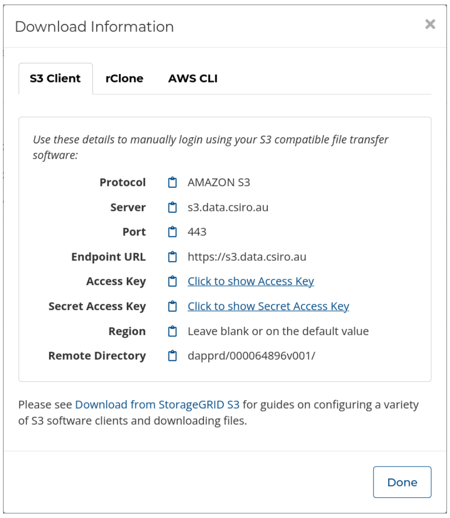

Figure 4: Example download pop-up from the CSIRO DAP.

3. The Wild-Places dataset must be downloaded from this link and placed in its own directory (outside of the CS-Wild-Places directory created in Step 2), which follows a similar process to the previous two steps using the CSIRO DAP. We provide instructions and scripts for post-processing and preparing the Wild-Places sequences on the HOTFormerLoc GitHub. Note that the default instructions will produce submaps for the post-processed and normalised format, but the `postprocess_wildplaces_ground.py` script supports other post-processing options via command-line arguments. Follow these instructions to post-process Wild-Places submaps and configure the directory structure of CS-Wild-Places.

4. Once Wild-Places is processed, the CS-Wild-Places directory should look similar to Fig. 2, with all sequences present for Karawatha and Venman. To reproduce the results reported in the paper, follow the instructions here to generate the training and evaluation pickles. Pre-trained model weights and instructions for using them are provided here.

References

[1] E. Griffiths, M. Haghighat, S. Denman, C. Fookes, and M. Ramezani, "HOTFormerLoc: Hierarchical Octree Transformer for Versatile Lidar Place Recognition Across Ground and Aerial Views," in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2025.

[2] J. Knights, K. Vidanapathirana, M. Ramezani, S. Sridharan, C. Fookes, and P. Moghadam, "Wild-Places: A Large-Scale Dataset for Lidar Place Recognition in Unstructured Natural Environments," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11322–11328, May 2023.