

WildScenes: A Benchmark for 2D and 3D Semantic Segmentation in Large-scale Natural Environments - Supplementary Material

Kavisha Vidanapathirana^{*,1,2}, Joshua Knights^{*,1,2}, Stephen Hausler^{*,1},
Mark Cox¹, Milad Ramezani¹, Jason Jooste¹, Ethan Griffiths^{1,2},
Shaheer Mohamed^{1,2}, Sridha Sridharan², Clinton Fookes², Peyman Moghadam^{1,2}

Abstract: Recent progress in semantic scene understanding has primarily been enabled by the availability of semantically annotated bi-modal (camera and lidar) datasets in urban environments. However, such annotated datasets are also needed for natural, unstructured environments to enable semantic perception for applications, including conservation, search and rescue, environment monitoring, and agricultural automation. Therefore, we introduce WildScenes, a bi-modal benchmark dataset consisting of multiple large-scale traversals in natural environments, including semantic annotations in high-resolution 2D images and dense 3D lidar point clouds, and accurate 6-DoF pose information. The data is (1) trajectory-centric with accurate localization and globally aligned point clouds, (2) calibrated and synchronized to support bi-modal inference, and (3) containing different natural environments over 6 months to support research on domain adaptation. Our 3D semantic labels are obtained via an efficient automated process that transfers the human-annotated 2D labels from multiple views into 3D point clouds, thus circumventing the need for expensive and time-consuming human annotation in 3D. We introduce benchmarks on 2D and 3D semantic segmentation and evaluate a variety of recent deep-learning techniques to demonstrate the challenges in semantic segmentation in natural environments. We propose train-val-test splits for standard benchmarks as well as domain adaptation benchmarks and utilize an automated split generation technique to ensure the balance of class label distributions. If you find this dataset helpful for your research, please cite the WildScenes paper [2].

1 How to Download WildScenes

This section provides a step-by-step guide to downloading the WildScenes Dataset. As the WildScenes dataset has a large total file size (96 GB), we recommend downloading it using an S3 client. These instructions differ depending on your OS (Operating System).

1.1 Linux

We recommend installing the command line application `reclone`, which can be performed using:

^{*}Equal Contribution

¹ Authors are with CSIRO Robotics, DATA61, CSIRO, Brisbane, QLD 4069, Australia. E-mails: *firstname.lastname@data61.csiro.au*

² Authors are with the research program in Signal Processing, Artificial Intelligence and Vision Technologies (SAIVT) at the Queensland University of Technology (QUT), Brisbane, Australia.

```
sudo apt install rclone
```

Then select the option “Download files via S3 Client” on the data portal website.

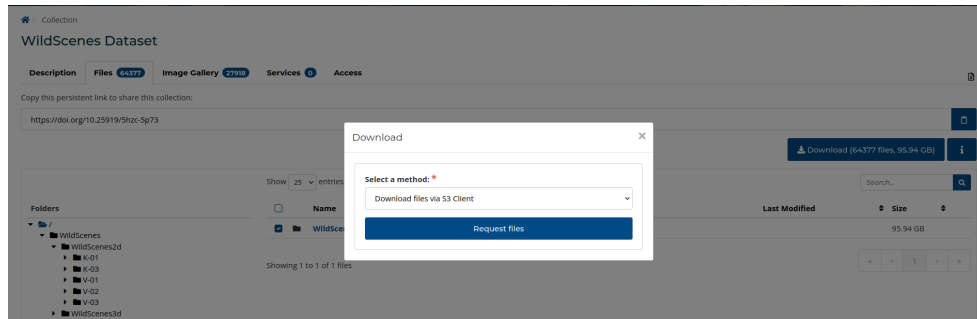


Figure 1: Popup box with selection to download files via S3 client

Then click the second tab on the next popup to use rclone.

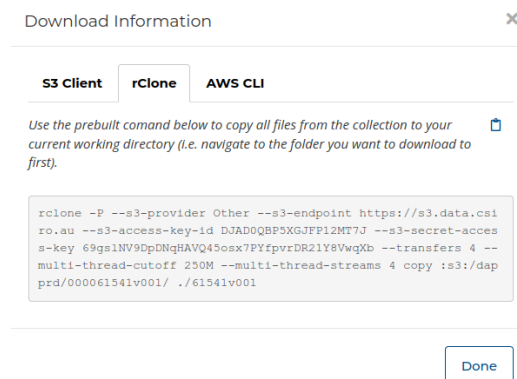


Figure 2: Popup box with command line instructions for terminal download

Then copy the provided text and paste in your Linux terminal (with rclone pre-installed). Download will commence automatically.

1.2 Windows/Mac

On Windows and Mac systems, both command line installation via rclone is supported, along with graphical S3 client software. On Windows, WinSCP could be used, on MacOS, Cyberduck could be used. Please navigate to the first tab shown in Figure 3, then use the details provided in this popup to login using your S3 compatible software.

For further assistance regarding downloading WildScenes, please refer to the following two web-sites:

- <https://research.csiro.au/dap/download/Download-SelectDownloadMethod>
- <https://research.csiro.au/dap/download/download-using-s3-compatible-software/>

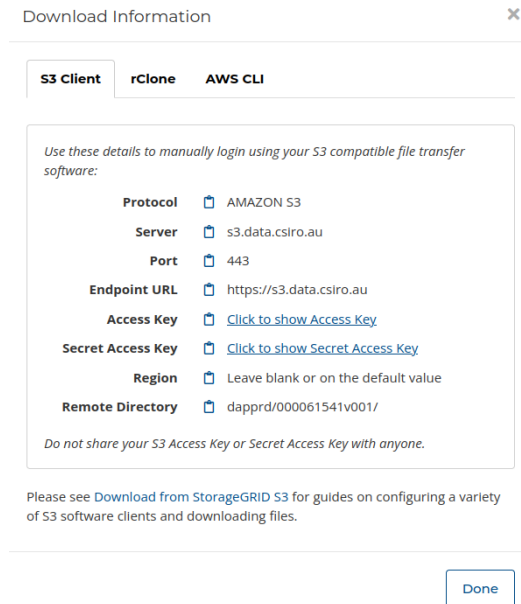


Figure 3: Popup box with details for using S3 clients

2 WildScenes Dataset Description

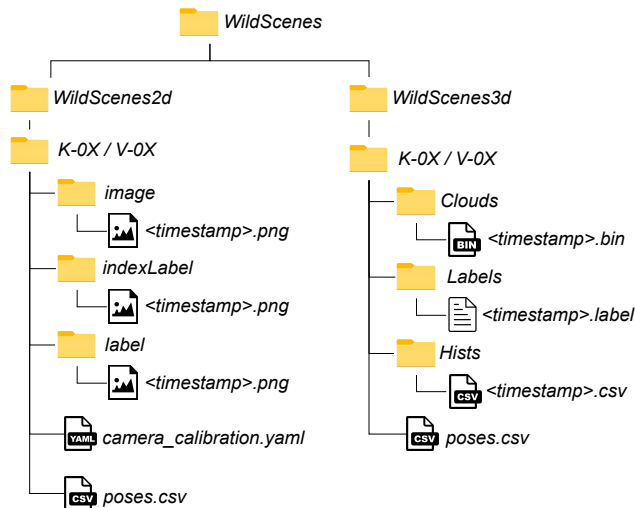


Figure 4: WildScenes Semantic Segmentation Dataset File Structure

The dataset release contains two components: the WildScenes semantic segmentation dataset, and a supplementary dataset which contains raw LiDAR point clouds. The raw point clouds are released to support applications beyond semantic segmentation, with further details provided in

Section 3. The remainder of Section 2 will focus on the semantic segmentation component.

The WildScenes semantic segmentation benchmark dataset contains annotated images and point clouds collected from two environments denoted as K and V. The file structure of the dataset is shown in Figure 4. Table 1 outlines the contents of each sequence and the respective file types, with further details provided in subsections below.

Table 1: Description of each file type in our dataset.

| File/Folder Name | Description |
|-------------------------|---|
| image | Sampled and rectified .png RGB images (2016×1512 resolution). |
| indexLabel | Raw annotated images in .png format. Pixel values are label index values (class indices [0-14] as assigned with classes sorted alphabetically by class name). |
| label (2D) | Palette annotated images in .png format (human readable label images). |
| Clouds | Point cloud submaps stored in .bin format and the same format as SemanticKITTI [1]. Point clouds are in the local frame of reference of the lidar sensor. |
| Labels (3D) | 3D labels are stored as .label files, again in the same format as SemanticKITTI. |
| Hists | .csv files containing the full histogram of 2D label observations for every 3D point. |
| poses.csv | .csv file containing timestamps and associated 6-DoF poses as calculated by SLAM. These poses are aligned such that all traverses have the same origin point in each environment (V and K). |
| camera_calibration.yaml | Sensor platform intrinsics and extrinsics, one .yaml file for each sequence of the dataset. |

2.1 3D Annotations

The 3D annotations consist of cloud files, storing the positional data (xyz) of the pointcloud, and label files, storing the annotated labels for every individual point in the pointcloud. Additionally, we provide .csv files storing the full distribution of label annotations produced per point during *LabelCloud*. Files are named by timestamp, which is the corresponding timestamp for the given pointcloud and the associated 6DoF pose can be acquired by the .csv pose files included in each sequence.

The positional data is stored as binary (.bin) files, in the same format as used in SemanticKITTI [1]. These files can be opened using the Python package *Numpy*, using the following line of code:

```
import numpy as np

points = np.fromfile(cloud, dtype=np.float32).reshape(-1,3)
```

Where a cloud is an individual *timestamp.bin* file, storing the xyz positional data of every point in the pointcloud.

The annotated labels are stored as .label files. These can be opened in the same manner as cloud files:

```
import numpy as np

labels = np.fromfile(labelspath, dtype=np.int32)
```

Labels files contain the integer index of the associated class label for every point in the pointcloud. The points and labels files are ordered the same and have a one-to-one alignment, in that the first index of points corresponds to the first index of labels etc. Index labels can also be mapped to their corresponding palette colors and human readable names. The index mapping is provided below:

```
METAINFO = {
    "classes": (
        "unlabelled",
        "bush",
        "dirt",
        "fence",
        "grass",
        "gravel",
        "log",
        "mud",
        "other-object",
        "other-terrain",
        "rock",
        "sky",
        "structure",
        "tree-foliage",
        "tree-trunk",
        "water",
    ),
    "palette": [
        (0, 0, 0),
        (230, 25, 75),
```

```

        (60, 180, 75),
        (0, 128, 128),
        (128, 128, 128),
        (145, 30, 180),
        (128, 128, 0),
        (255, 225, 25),
        (250, 190, 190),
        (70, 240, 240),
        (170, 255, 195),
        (0, 0, 128),
        (170, 110, 40),
        (210, 245, 60),
        (240, 50, 230),
        (0, 130, 200),
    ],
    "cidix": [
        255,
        0,
        1,
        2,
        3,
        4,
        5,
        6,
        7,
        8,
        9,
        10,
        11,
        12,
        13,
        14
    ]
}

```

2.2 2D Annotations

The 2D annotations consist of images, saved in the PNG format. We save annotations in two different styles: *indexLabel* and *label*. In *indexLabel*, annotations provide the integer index of the associated class label for every pixel in the RGB image. *Label* provides the same data, but mapped to their corresponding palette colors. The index mapping in 2D is provided below:

```

METAINFO = {
    "classes": (
        "unlabelled",
        "asphalt",
        "dirt",
        "mud",
        "water",
        "gravel",
        "other-terrain",
    )
}

```

```

        "tree-trunk",
        "tree-foliage",
        "bush",
        "fence",
        "structure",
        "pole",
        "vehicle",
        "rock",
        "log",
        "other-object",
        "sky",
        "grass",
    ),
    "palette": [
        (0, 0, 0),
        (230, 25, 75),
        (60, 180, 75),
        (255, 225, 25),
        (0, 130, 200),
        (145, 30, 180),
        (70, 240, 240),
        (240, 50, 230),
        (210, 245, 60),
        (230, 25, 75),
        (0, 128, 128),
        (170, 110, 40),
        (255, 250, 200),
        (128, 0, 0),
        (170, 255, 195),
        (128, 128, 0),
        (250, 190, 190),
        (0, 0, 128),
        (128, 128, 128),
    ],
    "cidx": [
        0,
        1,
        2,
        3,
        4,
        5,
        6,
        7,
        8,
        9,
        10,
        11,
        12,
        13,
        14,
        15,
        16,
    ]

```

```

    17,
    18
  ]
}

```

Note that the class list provided above is then filtered down to a smaller set of 15 classes for our benchmarks.

2.3 Pose Files

WildScenes also provides 6DoF poses for every image and every pointcloud (submap) in the dataset, containing the 6DoF poses in a quaternion format. Two `poses.csv` files are provided per sequence, one containing image poses and the other containing pointcloud poses. To associate a pointcloud/image with its given pose, use the timestamp in the filename of the submap to find its corresponding line in the `.csv`. All poses across all traverses are aligned and share a common origin point per traverse (one origin for environment V and another for K).

3 Supplementary Data: Raw LiDAR Point Clouds

To support applications beyond semantic segmentation, the dataset also includes the raw LiDAR point clouds. The 3D point clouds released as part of the WildScenes benchmark are semantic clouds, which means that every 3D point in a given submap is visible in at least one camera image frame. Since these point clouds are therefore subsets of the full 360 degree LiDAR scan, we also release the full 360 degree point clouds as a supplementary dataset. These point clouds include positional data per point (xyz) and both LiDAR intensity and return number. Furthermore, these submap point clouds are also sampled at identical timestamps as the 2D images and 3D semantic clouds from the main benchmark. These point clouds are stored in a `.ply` format and can be opened with any standard library for 3D point clouds including Open3D and also software such as CloudCompare. The file structure for this supplementary data is shown below:

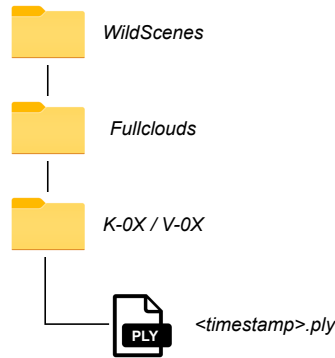


Figure 5: WildScenes Semantic Segmentation Dataset File Structure

References

- [1] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9297–9307, 2019.
- [2] Kavisha Vidanapathirana, Joshua Knights, Stephen Hausler, Mark Cox, Milad Ramezani, Jason Jooste, Ethan Griffiths, Shaheer Mohamed, Sridha Sridharan, Clinton Fookes, and Peyman Moghadam. Wildscenes: A benchmark for 2d and 3d semantic segmentation in large-scale natural environments. *The International Journal of Robotics Research*, 0(0):02783649241278369, 0.