The BigEarthNet Archive

The BigEarthNet archive was constructed by the Remote Sensing Image Analysis (RSiM) Group and the Database Systems and Information Management (DIMA) Group at the Technische Universität Berlin (TU Berlin). This work is supported by the European Research Council under the ERC Starting Grant BigEarth and by the German Ministry for Education and Research as Berlin Big Data Center BBDC.

BigEarthNet is a new large-scale Sentinel-2 benchmark archive, consisting of 590, 326 Sentinel-2 image patches. To construct BigEarthNet, 125 Sentinel-2 tiles acquired between June 2017 and May 2018 over the 10 countries (Austria, Belgium, Finland, Ireland, Kosovo, Lithuania, Luxembourg, Portugal, Serbia, Switzerland) of Europe were initially selected. All the tiles were atmospherically corrected by the Sentinel-2 Level 2A product generation and formatting tool (sen2cor). Then, they were divided into 590, 326 non-overlapping image patches. Each image patch was annotated by the multiple land-cover classes (i.e., multi-labels) that were provided from the CORINE Land Cover database of the year 2018 (CLC 2018).

If you use BigEarthNet, please cite:

G. Sumbul, M. Charfuelan, B. Demir, V. Markl, "BigEarthNet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding", IEEE International Geoscience and Remote Sensing Symposium, pp. 5901-5904, Yokohama, Japan, 2019.

Archive Directory Structure:

Each image patch in BigEarthNet has one directory under the archive root directory. GeoTIFF files for spectral bands, multi-labels and metadata of each patch will be in the corresponding patch directory. The directory structure of the archive is shown in following table. Different levels of the directory hierarchy are shown in different colors.

Directory Hierarchy	Description
<archive-root>/	<archive-root> path: BigEarthNet/
S2A_MSIL2A_20170717T113321_0_0/	Directory of the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_B01.tif	GeoTIFF file of band 1 for the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_B02.tif	GeoTIFF file of band 2 for the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_B03.tif	GeoTIFF file of band 3 for the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_B04.tif	GeoTIFF file of band 4 for the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_B05.tif	GeoTIFF file of band 5 for the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_B06.tif	GeoTIFF file of band 6 for the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_B07.tif	GeoTIFF file of band 7 for the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_B08.tif	GeoTIFF file of band 8 for the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_B8A.tif	GeoTIFF file of band 8A for the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_B09.tif	GeoTIFF file of band 9 for the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_B11.tif	GeoTIFF file of band 11 for the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_B12.tif	GeoTIFF file of band 12 for the patch S2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_0_labels_metadata.json	JSON file of multi-labels and metadata information for the patch 2A_MSIL2A_20170717T113321_0_0
S2A_MSIL2A_20170717T113321_0_1/	Directory of the patch S2A_MSIL2A_20170717T113321_0_1
...	
...	
S2A_MSIL2A_20170717T113321_1_0/	Directory of the patch S2A_MSIL2A_20170717T113321_1_0
S2A_MSIL2A_20170717T113321_1_1/	Directory of the patch S2A_MSIL2A_20170717T113321_1_1
...	
S2A_MSIL2A_20170717T113321_2_0/	Directory of the patch S2A_MSIL2A_20170717T113321_2_0
...	

Naming Conventions:

The compact naming convention for each patch directory is defined as follows:

- **<sentinel-mission-id>_MSIL2A_<YYYYMMDD>T<HHMMSS>_<horizontal-patch-order>_<vertical-patch-order>**

  Components of each folder name are defined as follows:

- **<sentinel-mission-id>** is the Sentinel-2 mission ID that can be either S2A or S2B.
- **MSIL2A** denotes the 2A product level of Sentinel-2 tiles.
- **<YYYYMMDD>** is the acquisition date of a Sentinel-2 tile including year, month and day information. For instance, 20170717 denotes 'July17th, 2017'.
- **<HHMMSS>** is the acquisition time of a Sentinel-2 tile including hour, minute and second information. For the time convention, 24-hour clock format is used.
- **<horizontal-patch-order>** identifies the horizontal order of the patch in the tile from which the patch is extracted. This number starts with 0.
- **<vertical-patch-order>** identifies the vertical order of the patch in the tile from which the patch is extracted. This number starts with 0.

File Formats:

- Each spectral band is stored in a separate GeoTIFF file as a georeferenced raster image. Names of these files are defined by adding band names together with the tif extension (_B01.tif, _B02.tif etc.) to patch folder names.

- Multi-labels and metadata of each patch are stored in a JSON file with a name extension (_labels_metadata.json) to the patch folder name. Name-value pairs of this file with explanations are as follows:

{

"labels" : [“<label>”, “<label>”,…,”<label>”] (Multiple class names in an array structure),

"tile_source" : “<1C_Level_Product_Name>” (Original unprocessed tile, which can be obtained from Copernicus Open Access Hub, product name),

"acquisition_date" : "<YYYY-MM-DD HH:MM:SS>" (Acquisition date and time of the corresponding tile)

"coordinates" : **{**

"ulx" : <upper_left_x>, **"uly"** : <upper_left_y>, (Upper left corner coordinate of the patch)

"lrx" : <lower_right_x>, **"lry"** : <lower_right_y>, (Lower right corner coordinate of the patch)

},

"projection" : "<wkt_projection>" (Projection for the patch coordinates in Well-Known Text format (WKT))

}