

SEN12-FLOOD Dataset Documentation

Name:

SEN12-FLOOD: A SAR and Multispectral Dataset for Flood Detection

Publication Date:

September 18, 2020

Publication resource:

Available (primary repository) from: <https://iee-dataport.org/open-access/sen12-flood-sar-and-multispectral-dataset-flood-detection>

DOI: <https://dx.doi.org/10.21227/w6xz-s898>

Version:

1.0

Description:

SEN12-FLOOD is a set of multimodal (SAR + multispectral) satellite image time-series for flood classification.

The observed areas correspond to 337 locations (cities and their surroundings) in West and South-East Africa, Middle-East, and Australia where a flood event occurred during the considered period. The period of acquisition goes from December 2018 to May 2019.

For each location, the following data are provided:

- Time series of Sentinel-2 multispectral images. These images are composed of 12 bands, at 10m ground-sampling distance and are provided with Level 2A atmospheric correction.
- Time series of Sentinel-1 Synthetic Aperture Radar (SAR) images. The images are provided with radiometric calibration and range doppler terrain correction based on the SRTM digital elevation model. For one acquisition, two raster images are available corresponding to the polarimetry channels VV and VH.
- Time series of binary labels for each image / date: flood or no flood.

The original dataset was split into 262 sequences for the train and 68 sequences for the test.

The authors of [1] created the first dataset with Sentinel-2 images and flood labels. The authors of [2] added the Sentinel-1 images and corresponding, new flood labels.

Methodology:

The city-centered satellite sequences provided by the Media-Eval 2019 Multimedia Satellite task [1] give access to series of multispectral Sentinel 2 images (<http://www.multimediaeval.org/mediaeval2019/multimediasatellite/>).

In [2] we proposed a new dataset corresponding to the Sentinel 1 sequences for the same areas and periods. However, since SAR is independent of cloud cover, more SAR images are retrieved for the same time period, leading to a higher sampling rate. This SAR dataset is composed of roughly

two times more images than the optical one. The Sentinel 1 images were downloaded from the Scientific ESA hub website (<https://scihub.copernicus.eu/>).

Each image has a binary label specifying whether a flood event is visible or not in the observed area. The labels have been provided by the original MediaEval 2019 dataset and were obtained from the Copernicus Emergency Management Service (<https://emergency.copernicus.eu/>).

To leverage both SAR and optical modalities, we merged the MediaEval dataset and our own in the new SEN12-FLOOD dataset.

After discussions with the MediaEval organizers (namely Benjamin Bischke), we agreed to release the new dataset publicly to foster research and development for flood mapping and natural disaster recovery.

Class Definitions:

The task addressed here is entire image classification. For each image / date in the time series, one single binary label is assigned: flood or no flood.

A flood event is occurring in 40% of the optical Sentinel 2 images and in 47% of the SAR Sentinel 1 images. As in the MediaEval dataset, once a flood occurred in a sequence, all the subsequent images are labelled as flooded which corresponds to the hypothesis that the surface still presents particular modifications after the event.

Coordinate Reference System:

Sentinel-2 image chips are stored in UTM / WGS84 coordinates. Sentinel-1 image chips are stored using cartesian lat/lon (epsg 4326).

File Name Structure:

All images are stored in raster format (GeoTIFF files).

Two json files, S1list.json and S2list.json are provided to describe respectively the Sentinel-1 and Sentinel-2 images. The keys are the total number of images in the sequence, the folder name, the geography of the observed area, and the description of each image in the series. The SAR images description contains also the URLs to download the images. Each image is described by its acquisition date, its label (FLOODING: boolean), a boolean (FULL-DATA-COVERAGE: boolean) indicating if the area is fully or partially imaged, and the file prefix. For SAR images the orbit (ASCENDING or DESCENDING) is also indicated.

Spatial Extent:

The observed areas correspond to 337 locations (cities and their surroundings) in West and South-East Africa, Middle-East, and Australia where a flood event occurred.

Temporal Extent:

The images and the flood labels correspond to the period going from 01/12/2018 to 31/05/2019.

Licence:

The dataset contains modified Copernicus data 2018-2019.

Copernicus Sentinel data are available under the Sentinel Terms and Conditions:

https://scihub.copernicus.eu/twiki/pub/SciHubWebPortal/TermsConditions/TC_Sentinel_Data_31072014.pdf

In particular, the User Rights article grant the User with Open-Access style permissions (use, modification, sharing, publication and distribution).

Citation:

[1] The Multimedia Satellite Task at MediaEval2019, Bischke, B., Helber, P., Schulze, C., Srinivasan, V., Dengel, A., Borth, D., 2019, In Proc. of the MediaEval 2019 Workshop

[2] Flood Detection in Time Series of Optical and SAR Images, C. Rambour, N. Audebert, E. Koeniguer, B. Le Saux, and M. Datcu, ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2020, 1343--1346

Contact:

Bertrand.Le.Saux@esa.int; clement.rambour@lecnam.net