

TimeSen2Crop: a Million Labeled Samples Dataset of Sentinel 2 Image Time Series for Crop Type Classification

April 2021

Description

This benchmark dataset has been developed in the framework of the ExtremeEarth project, which received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 825258.

TimeSen2Crop is a pixel based dataset made up of more than 1 million samples of Sentinel 2 Time Series associated to 16 crop types, during an agronomic year ranging from September 2017 to August 2018. The dataset contains atmospherically corrected samples, as well as information related to snow, clouds and shadows. Table 1 summarizes the number of training samples available for each tile and each crop type. TimeSen2Crop can be downloaded at the following link [TimeSen2Crop](#). For an in-depth analysis about the performances and the procedures followed to obtain the the proposed dataset, please refer to [1].

	Tiles															
	2019_33UVP	32TNT	32TPT	32TQT	33TUM	33TUN	33TVM	33TVN	33TWM	33TWN	33TXN	33UUP	33UVP	33UWP	33UWQ	33UXP
Legumes	853	0	1	0	1	0	48	52	114	558	293	1042	2031	1460	963	535
Grassland	20061	20001	20001	20001	20003	20001	20017	20001	20013	20033	9602	20037	15080	15058	15095	8259
Maize	20001	1188	1684	1308	2872	486	14131	2257	20001	20001	6930	20001	15001	15001	8453	15001
Potato	2910	66	458	111	144	101	232	264	165	559	555	270	4015	11121	5274	4433
Sunflower	151	0	0	1	3	0	113	0	56	1313	1837	100	240	6672	1350	10951
Soy	9750	0	0	7	34	8	2520	113	5308	14011	7128	3750	10712	6336	468	10739
Winter Barley	20001	23	33	76	434	88	2028	275	2849	9107	2460	12194	15001	15001	2586	11905
Winter Caraway	202	0	0	0	0	0	0	0	0	19	0	90	577	303	183	98
Rye	7503	6	52	8	76	19	477	394	119	3807	1750	738	9701	12093	10712	6239
Rapeseed	3923	0	0	2	0	0	30	5	94	2168	4005	3211	5086	9640	5771	7966
Beet	3573	0	7	2	1	0	8	62	62	766	1556	344	4212	12478	2218	8775
Spring Cereals	8111	6	229	24	474	287	1857	828	376	4819	2931	3193	11987	19912	14896	15423
Winter Wheat	20001	2	47	67	111	82	1659	495	3734	18490	14865	12770	15001	15001	15001	15001
Winter Triticale	12342	15	82	64	551	88	1810	886	1428	4369	629	5292	14363	12559	9394	2576
Permanent Plantation	210	62	97	14	7	1	62	4	3810	6348	4888	11	411	14428	3810	12149
Other Crops	10001	432	1841	1399	1306	10001	301	10001	657	10001	123	10001	10001	10001	229	418
Total	139593	21801	24532	23084	26017	31162	45293	35637	58786	116369	59552	93044	133419	177064	96403	130468

Files organization

The dataset is organized hierarchically, as shown in Fig. 1. The data are organized per Sentinel 2 tiles, in 16 folders (15 tiles plus one tile for the 2019 agronomic year). Each Sentinel 2 folder contains 16 sub-folders where samples for each crop type are stored. The Sentinel 2 folder contains also a csv file with the acquisition dates for the samples in that particular tile, ordered from the oldest acquisition to the newest. Inside the crop type sub-folders, each labeled sample is stored in a csv file containing the multispectral temporal signature. The csv files provides a matrix where each row indicates the acquisition date and each column indicates the spectral band associated, namely blue (B2), green (B3), red (B4), the four vegetation red edge (B5, B6, B7 and B8A) and the two short wave infrared (SWIR) (B11 and B12). The last column contains information regarding the condition of the pixel: clear (0), cloud (1), shadow (2) or snow (3).

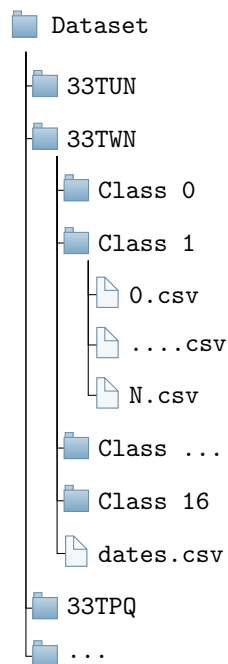


Figure 1: File structure of the dataset.

References

- [1] Giulio Weikmann, Claudia Paris, and Lorenzo Bruzzone. “TimeSen2Crop: a Million Labeled Samples Dataset of Sentinel 2 Image Time Series for Crop Type Classification”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2021), pp. 1–1. DOI: [10.1109/JSTARS.2021.3073965](https://doi.org/10.1109/JSTARS.2021.3073965).