SEN12MS – A CURATED DATASET OF GEOREFERENCED MULTI-SPECTRAL SENTINEL-1/2 IMAGERY FOR DEEP LEARNING AND DATA FUSION

M. Schmitt[1], L. H. Hughes[1], C. Qiu[1], X. X. Zhu[1,2]

[1] Signal Processing in Earth Observation, Technical University of Munich, Munich, Germany
- (m.schmitt, lloyd.hughes, chunping.qiu)@tum.de
[2] Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Wessling
- xiaoxiang.zhu@dlr.de

KEY WORDS: Data Fusion, Dataset, Machine Learning, Remote Sensing, Multi-Spectral Imagery, Synthetic Aperture Radar (SAR), Optical Remote Sensing, Sentinel-1, Sentinel-2, Deep Learning

ABSTRACT:

The availability of curated large-scale training data is a crucial factor for the development of well-generalizing deep learning methods for the extraction of geoinformation from multi-sensor remote sensing imagery. While quite some datasets have already been published by the community, most of them suffer from rather strong limitations, e.g. regarding spatial coverage, diversity or simply number of available samples. Exploiting the freely available data acquired by the Sentinel satellites of the Copernicus program implemented by the European Space Agency, as well as the cloud computing facilities of Google Earth Engine, we provide a dataset consisting of 180,662 triplets of dual-pol synthetic aperture radar (SAR) image patches, multi-spectral Sentinel-2 image patches, and MODIS land cover maps. With all patches being fully georeferenced at a 10 m ground sampling distance and covering all inhabited continents during all meteorological seasons, we expect the dataset to support the community in developing sophisticated deep learning-based approaches for common tasks such as scene classification or semantic segmentation for land cover mapping.

1. INTRODUCTION

The availability of curated annotated datasets is of crucial importance for the development of machine learning models for information retrieval from remote sensing data. While classic *shallow* learning approaches could easily be trained on comparably small datasets such as, e.g., the famous Indian Pines scene (Baumgardner et al., 2015), modern *deep* learning requires large-scale data to reach the desired generalization performance (Zhu et al., 2017). However, computer vision usually deals with conventional photographs of everyday objects, whereas remote sensing data is more versatile and much more difficult to interpret. Therefore, massive databases of labeled imagery such as ImageNet (Deng et al., 2009) do not yet exist in the remote sensing domain, although there have been first steps into that direction; a certainly non-exhaustive overview of existing scientific datasets of annotated remote sensing imagery can be found in Table 1. Additional datasets, which were mostly provided in the frame of machine learning competitions and are not described and discussed in scientific papers, can be found in a the private link list of (Rieke, 2019).

In order to support deep learning-related research in the field of remote sensing, we have published the *SEN1-2* dataset in 2018, which is comprised of about 280,000 pairs of corresponding Sentinel-1 SAR and Sentinel-2 optical images (Schmitt et al., 2018). Since *SEN1-2* was mainly intended for bridging the gap between classical computer vision problems and remote sensing, e.g. image-to-image translation tasks, the provided data were strongly simplified: For Sentinel-1, we just provided vertically polarized (VV) imagery in dB scale, and for Sentinel-2 we reduced the original multi-spectral data tensors to RGB images with adjusted histograms, whereas none of the images came with any form of geolocation information. Based on feedback from the community, with this paper, we now publish a follow-on version of the dataset, which is designed to suit the needs of the remote sensing community. *SEN12MS* contains full multi-spectral information in geocoded imagery and is described in details throughout the remainder of this paper.

2. THE DATA BASIS

We exploit freely available satellite(-derived) data to form the basis of the dataset. On the one hand, we make use of SAR and multi-spectral imagery provided by Sentinel-1 and Sentinel-2, respectively. On the other hand, we add land cover information derived from observations acquired by the MODIS system. Details of the three basic data sources are provided in the following.

2.1 Sentinel-1

The Sentinel-1 mission (Torres et al., 2012) consists of currently two polar-orbiting satellites, equipped with C-band SAR sensors, which enables them to acquire imagery regardless of the weather. Sentinel-1 works in a pre-programmed operation mode to avoid conflicts and to produce a consistent long-term data archive built for applications based on long time series. Depending on which SAR imaging mode is used, resolutions down to 5 m with a wide coverage of up to 400 km can be achieved. Furthermore, Sentinel-1 provides dual polarization capabilities and very short revisit times of about 1 week at the equator. Since highly precise spacecraft positions and attitudes are combined with the high accuracy of the range-based SAR imaging principle, Sentinel-1 images come with high out-of-the-box geolocation accuracy (Schubert et al., 2015).

For the Sentinel-1 images in the *SEN12MS* dataset, again ground-range-detected (GRD) products acquired in the most frequently available interferometric wide swath (IW) mode were used. These images contain the σ^0 backscatter coefficient in dB scale

https://doi.org/10.5194/isprs-annals-IV-2-W7-153-2019 | © Authors 2019. CC BY 4.0 License.

Dataset	Nr. of Images	Image Size	Data Source	Description	Reference
UC Merced Land Use Dataset	2,100	256 × 256	Color aerial images	21 balanced land use classes	(Yang and Newsam, 2010)
SAT-4	500,000	28 × 28	Color aerial imagery (R, G, B, NIR)	4 agricultural classes over continental USA	(Basu et al., 2015)
SAT-6	405,000	28 × 28	Color aerial imagery (R, G, B, NIR)	6 land cover classes over continental USA	(Basu et al., 2015)
Brazilian Coffee Scenes Dataset	51,004	64 × 64	SPOT	2 imbalanced classes describing non-coffee and coffee	(Penatti et al., 2015)
USGS SIRI-WHU	1	10,000 × 9,000	Color aerial image	4 classes over Montgomery County, Ohio, USA	(Zhong et al., 2015)
Brazilian Cerrado-Savanna Dataset	1,311	64 × 64	RapidEye (G, R, NIR)	4 imbalanced classes describing Cerrado-Savanna vegetation	(Nogueira et al., 2016)
SIRI-WHU	200	200 × 200	Google Earth	12 classes describing urban areas in China	(Zhao et al., 2016)
Inria Aerial Image Labeling Dataset	360	1,500 × 1,500	Color aerial imagery	2 classes (building / not building) for 10 cities in Austria and USA	(Maggiori et al., 2017)
2017 IEEE GRSS Data Fusion Contest	57	from 447 × 377 to 1,461 × 1,222	Sentinel-2, Landsat, OpenStreetMap	17 local climate zone classes	(Yokoya et al., 2018)
DeepGlobe – Road Extraction	8,570	1,024 × 1,024	Worldview-2/-3, GeoEye-1 (R, G, B)	1 target class: roads over India and Thailand	(Demir et al., 2018)
DeepGlobe – Building Detection	24,586	650 × 650	Worldview-3	1 target class: buildings in Las Vegas, Shanghai, Paris, Karthoum	(Demir et al., 2018)
DeepGlobe – Land Cover Classification	1,146	2,448 × 2,448	Worldview-2/-3, GeoEye-1 (R, G, B)	7 land cover classes	(Demir et al., 2018)
DOTA	2,806	800 × 800 to 4,000 × 4,000	Color aerial images	188,282 instances of 15 object classes, each labeled by a quadriliteral	(Xia et al., 2018)
EuroSAT	27,000	64 × 64	Sentinel-2	10 classes	(Helber et al., 2018)
SEN1-2	564,768	256 × 256	Sentinel-1 and Sentinel-2	corresponding pairs of SAR (single-pol intensity) and optical (RGB) image pairs without annotations	(Schmitt et al., 2018)
38-Cloud	17,601	384 × 384	Landsat 8 (R, G, B, NIR)	single target class: clouds	(Mohajerani and Saeedi, 2019)
BigEarthNet	590,326	up to 120 × 120	Sentinel-2	43 Corine Land Cover classes over Europe	(Sumbul et al., 2019)
SEN12MS	541,986	256 × 256	Sentinel-1, Sentinel-2, MODIS Land Cover	globally distributed; MODIS Land Cover maps can either be used as labels or auxiliary data	*this paper*

Table 1. Non-exhaustive list of curated datasets for deep learning in remote sensing.

for every pixel at a pixel spacing of 5 m in azimuth and 20 m in range. In order to exploit the full potential of Sentinel-1 data, *SEN12MS* contains both VV and VH polarized images. For precise ortho-rectification, restituted orbit information was combined with the 30 m-SRTM-DEM or the ASTER DEM for high latitude regions where SRTM is not available. As for the *SEN1-2* dataset, we intend to leave any further pre-processing, e.g. speckle filtering, to the end user and do not manipulate the data any further.

2.2 Sentinel-2

The Sentinel-2 mission (Drusch et al., 2012) currently comprises two identical polar-orbiting satellites in the same orbit, phased at 180° to each other. The mission is meant to provide continuity for multi-spectral imagery of the SPOT and LANDSAT kind, which have provided information about the land surfaces of our Earth for many decades. With its wide swath width of up to 290 km and its high revisit time of 5 days at the equator (based on two satellites) under cloud-free conditions, the Sentinel-2 mission is specifically well-suited to vegetation monitoring within the growing season.

For the *SEN12MS* dataset, we provide the full multi-spectral image cubes as extracted from the original, precisely georeferenced Sentinel-2 granules. The only manipulation we carried out was to implement a sophisticated mosaicking workflow to avoid the download of cloud-affected images (cf. Section 3.1).

2.3 MODIS Land Cover

MODIS (the Moderate Resolution Imaging Spectroradiometer) is the main instrument on board of the Terra and Aqua satellites. Terra's orbit around the Earth is timed so that it passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon. Terra MODIS and Aqua MODIS acquisitions cover the whole Earth with an approximately daily revisit frequency – at a band-dependent resolution of 250 m to 1000 m. Based on calibrated MODIS reflectance data, hierarchical classification following the land cover classification system (LCCS) scheme, and sophisticated post-processing for class-specific refinement incorporating prior knowledge, auxiliary information and temporal regularization based on a Markov random field, annually updated global land cover maps for the years 2001–2016 are provided as MCD12Q1 V6 dataset at a ground sampling distance of 500 m (Sulla-Menashe et al., 2019). To add land cover information to the Sentinel-1/Sentinel-2 patch-pairs constituting the core of the *SEN12MS* dataset, we add four-band MODIS land cover patches created from 2016 data at an upsampled pixel spacing of 10 m. The first of the provided bands contains land cover following the International Geosphere-Biosphere Programme (IGBP) classification scheme (Loveland and Belward, 1997), while the remaining bands contain the LCCS land cover layer, the LCCS land use layer, and the LCCS surface hydrology layer (Di Gregorio, 2005). The schemes' classes are listed in Table 2. According to (Sulla-Menashe et al., 2019), the overall accuracies of the layers are about 67% (IGBP), 74% (LCCS land cover), 81% (LCCS land use), and 87% (LCCS surface hydrology), respectively. This should be kept in mind when using the land cover data as labels for training scene classification or semantic segmentation models, as these accuracies will constitute the upper bound of actually achievable predictive power – even if validation accuracies of 100% are reached. If the land cover information is not utilized as annotation, but as auxiliary data source, similar caution should be had.

3. GOOGLE EARTH ENGINE FOR DATA PREPARATION

As for the *SEN1-2* dataset, we have again utilized Google Earth Engine (Gorelick et al., 2017) to generate a large-scale dataset of corresponding multi-sensor remote sensing image patches. While we basically used the same pipeline as described in (Schmitt et al., 2018), including the random sampling of ROIs for the meteorological seasons of the northern hemisphere, we added a more sophisticated mosaicking workflow for the generation of cloud-free short-term Sentinel-2 mosaics.

3.1 Mosaicking of Cloud-Free Sentinel-2 Images

The general mosaicking workflow to produce cloud-free Sentinel-2 images for a given region of interest (ROI) and a specified time period is depicted in Figure 1. In essence, it consists of three main modules, which are carried out for every ROI. These ROIs result from two uniform random samplings over the landmasses of the Earth and the urban areas across the globe, respectively.

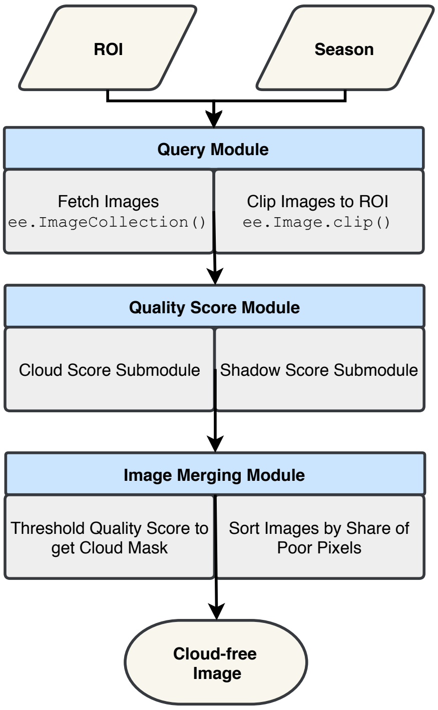

Figure 1. Workflow of the GEE-based procedure for cloud-free Sentinel-2 image generation presented in (Schmitt et al., 2019).

While the procedure is described in detail in (Schmitt et al., 2019), a short summary of the three modules is as follows:

(1) The *Query Module* for loading images from the catalogue. In this module, for the specified ROI all Sentinel-2 images available for a specified time period are selected.

(2) The *Quality Score Module* for the calculation of a quality score for each image. In this module, every pixel of each Sentinel-2 image is assigned a score that considers the likelihood it is affected either by clouds or by shadow.

(3) The *Image Merging Module* for mosaicking of the selected images based on the meta-information generated in the preceding modules. First, the quality scores are thresholded to determine cloud and shadow masks for each image. Afterwards, the images are sorted by their amount of poor pixels. The best images are finally merged into a cloud-free mosaic.

Since the Sentinel-1 images and the MODIS land cover data are not affected by clouds, in these cases no complicated mosaicking processes are required and the data are simply exported in a straight-forward manner later on.

3.2 Data Export

For *SEN12MS* we have utilized the same random ROIs as for *SEN1-2*. The same holds for the meteorological seasons as defined for the northern hemisphere. After preparation of the Sentinel-1 images and the cloud-free Sentinel-2 mosaics for every ROI and season, we export them together with the land cover

Class	IGBP value	LCCS LC value	LCCS LU value	LCCS SH value
Evergreen Needleleaf Forests	1	11	-	-
Evergreen Broadleaf Forests	2	12	-	-
Deciduous Needleleaf Forests	3	13	-	-
Deciduous Broadleaf Forests	4	14	-	-
Mixed Broadleaf/Needleleaf Forests	-	15	-	-
Mixed Broadleaf Evergreen/Deciduous Forests	-	16	-	-
Mixed Forests	5	-	-	-
Dense Forests	-	-	10	10
Open Forests	-	21	20	20
Sparse Forests	-	22	-	-
Natural Herbaceous	-	-	30	-
Dense Herbaceous	-	31	-	-
Sparse Herbaceous	-	32	-	-
Shrublands	-	-	40	40
Closed (Dense) Shrublands	6	41	-	-
Open (Sparse) Shrublands	7	43	-	-
Shrubland/Grassland Mosaics	-	42	-	-
Woody Savannas	8	-	-	-
Savannas	9	-	-	-
Grasslands	10	-	-	30
Permanent Wetlands	11	-	-	-
Woody Wetlands	-	-	-	27
Herbaceous Wetlands	-	-	-	50
Herbaceous Croplands	-	-	36	-
Croplands	12	-	-	-
Urban and Built-Up Lands	13	-	9	-
Cropland/Natural Vegetation Mosaics	14	-	-	-
Forest/Cropland Mosaics	-	-	25	-
Natural Herbaceous/Croplands Mosaics	-	-	35	-
Tundra	-	-	-	51
Permanent Snow and Ice	15	2	2	2
Barren	16	1	1	1
Water Bodies	17	3	3	3

Table 2. MODIS land cover classes as represented by four different schemes: IGBP, LCCS land cover, LCCS land use, LCCS surface hydrology. The *NoData* value is set to be 255 (Sulla-Menashe and Friedl, 2018).

data at a scale of 10 m in the form of GeoTiffs. In this context, it has to be noted that the 10 m scale is defined at the equator by Google Earth Engine, which corresponds to an angular resolution of $0.0001°$. This angular resolution leads to significantly smaller pixel widths for regions with latitudes deviating from $0°$. We therefore used GDAL (Warmerdam, 2008) to transform all exported data from the WGS84 lat/lon georeference to their local UTM coordinate representation, while resampling to actually square pixels of $10\,\text{m}\times10\,\text{m}$.

3.3 Data Curation

After the export of the data from the GEE servers to local storage, we followed an inspection protocol similar to the one proposed in our previous work (Schmitt et al., 2018): First, each triplet of full scene images was converted to a visually perceivable format (gray-scale images for Sentinel-1 and MODIS Land Cover, RGB images for Sentinel-2) and displayed to a remote sensing expert. If either of the three images contained very large no-data areas, large non-detected clouds, or strong artifacts resulting from the cloud-adaptive mosaicking, the triplet was discarded. After this first inspection, only 252 out of the originally downloaded 600 scenes were kept in the dataset. These remaining scenes were then tiled into patches of 256×256 pixels in size. Again, we have implemented a stride of 128 pixels, resulting in an overlap between adjacent patches of 50%. We think, a 50% overlap is the ideal trade-off between patch independence and maximization of the number of samples. After the tiling, 216,596 patch triplets were available for a second inspection. In this second inspection, all patches were again visually inspected by remote sensing experts in order to avoid patches containing artefacts or distortions, e.g. no data areas, clouds, or jet streams. Some examples for patch types that were discarded in this step are displayed in Figure 2. After this final inspection step, a total of 180,662 patch triplets remained, which comprise the final *SEN12MS* dataset. The locations of the final ROI scene locations are displayed in Figure 3.

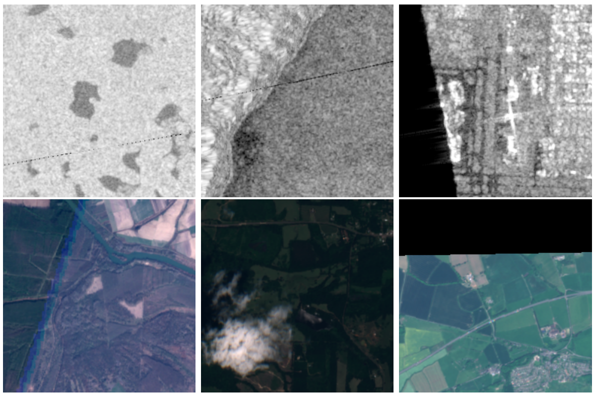

Figure 2. Examples for manually removed SAR (top row) and optical patches (bottom) row).

4. THE SEN12MS DATASET

The final *SEN12MS* dataset contains 180,662 patch triplets (Sentinel-1 dual-pol SAR, Sentinel-2 multi-spectral, MODIS land cover) in the form of multi-channel GeoTiff images and requires 421.3 GiB of storage. Some patch triplet examples are

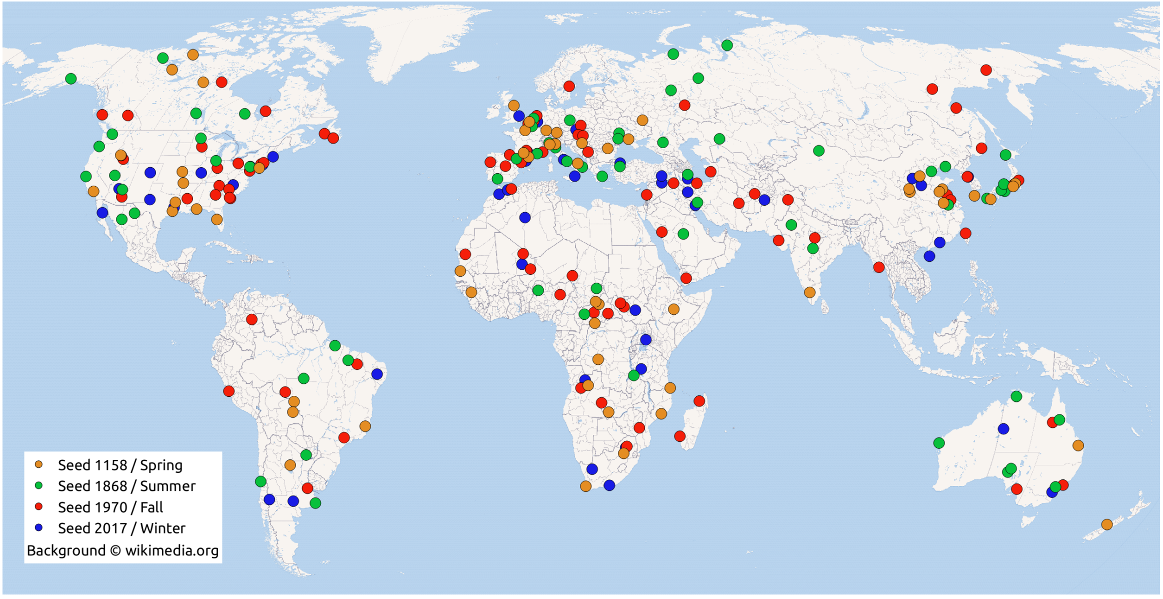

Figure 3. Final distribution of the ROIs over all inhabited land masses of the Earth.

shown in Figure 4 to give an impression of the rich and versatile information contained in the dataset.

4.1 Structure of the Final Dataset

As mentioned in Section 3.1, the dataset is based on randomly sampled regions of interest, resulting from four different seed values: `1158`, `1868`, `1970`, and `2017`. These four different seed values are related to the four meteorological seasons defined for the northern hemisphere: winter (1 December 2016 to 28 February 2017), spring (1 March 2017 to 30 May 2017), summer (1 June 2017 to 31 August 2017), and fall (1 September 2017 to 30 November 2017). This leads to a tree-like structure of the dataset into four branches: `ROIs1158_spring`, `ROIs1868_summer`, `ROIs1970_fall`, and `ROIs2017_winter`. Each of those branches again is divided into several sub-branches corresponding to the individual ROIs (or scenes, respectively) derived from the corresponding random number seed. The full dataset structure is depicted in Figure 5.

The Sentinel-1 data can be recognized by the abbreviation `s1`, the Sentinel-2 data by `s2`, and the MODIS land cover data by `lc`; the individual patches can be identified by the token `pXXX` where `XXX` denotes a unique identifier number per patch. Thus, the file naming convention follows the following scheme: `ROIsSSSS_SEASON_DD_pXXX.tif`, where `SSSS` denotes the seed value, `SEASON` denotes the meteorological season as defined for the northern hemisphere, `DD` denotes the data identifier, and `XXX` denotes the patch identifier.

Of course, we are aware that the seasonal structuring of the dataset is only of little semantic worth since we have taken the seasons of the northern hemisphere as a reference. To allow end-users a sub-structuring of the dataset taking semantically meaningful seasons into account, we provide the file `seasons.csv` with the metadata of the dataset. It declares, which scenes actually were acquired in spring, summer, winter, and fall from a climatic point of view.

While we forgo to define a fixed train/test split, we think this can easily be achieved by end-users considering their individual needs: Deterministic splits into disjunct training and test sets can

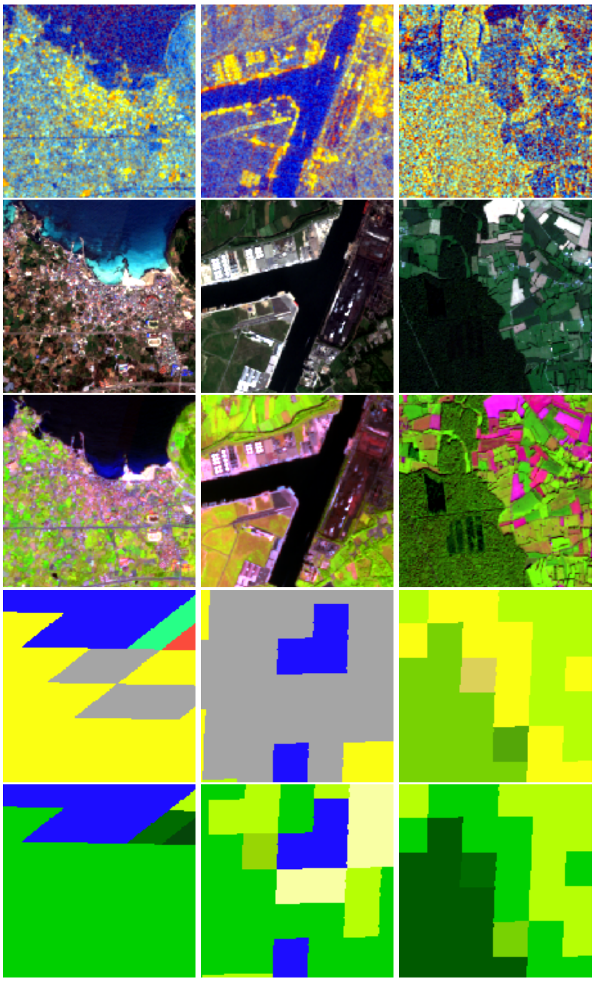

Figure 4. Images extracted from 3 example patch triplets. Each column shows (from top to bottom): False color Sentinel-1 SAR (R: VV, G: VH, B: VV/VH), Sentinel-2 RGB, Sentinel-2 SWIR, IGBP Land cover, LCCS Land cover. Note that while the GSD of all patches is upsampled to 10 m, the actual resolution varies from 10 m (Sentinel-2 RGB) to 500 m (land cover).

https://doi.org/10.5194/isprs-annals-IV-2-W7-153-2019 | © Authors 2019. CC BY 4.0 License.

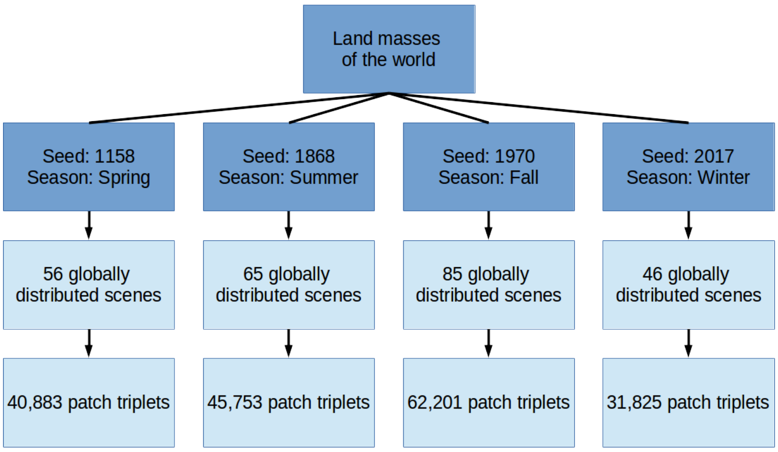
Land masses of the world
Seed: 1158 Season: Spring
Seed: 1868 Season: Summer
Seed: 1970 Season: Fall
Seed: 2017 Season: Winter
56 globally distributed scenes
65 globally distributed scenes
85 globally distributed scenes
46 globally distributed scenes
40,883 patch triplets
45,753 patch triplets
62,201 patch triplets
31,825 patch triplets

Figure 5. Tree structure of the final dataset.

be achieved via the meteorological seasons, or via the individual ROIs.

4.2 Dataset Availability

The *SEN12MS* dataset is shared under the open access license CC-BY and available for download at a persistent link provided by the library of the Technical University of Munich (TUM): `https://mediatum.ub.tum.de/1474000`. This paper must be cited when the dataset is used for research purposes.

5. APPLICATION TO LAND COVER MAPPING

In order to provide an example for the usefulness of the dataset with regard to the development of land cover classification solutions, we have trained two state-of-the-art deep convolutional neural network architectures for predicting LCCS land use classes (cf. Table 3). The first network, ResNet-110 (He et al., 2016), was designed for image classification, i.e. for assigning a single class label to the input image. For the presented baseline experiment, we cut images of 64 × 64 pixels from the Sentinel-2 samples from the *summer* subset, and used the following ten bands as channel information: B2 (Blue), B3 (Green), B4 (Red), B8 (Near-infrared), B5 (Red Edge 1), B6 (Red Edge 2), B7 (Red Edge 3), B8a (Red Edge 4), B11 (Short-wavelength infrared 1), and B12 (Short-wavelength infrared 2). We then used the majority LCCS land use class from each of the 64 × 64 patches as scene label for that patch. Due to the unconventional multi-channel configuration of the data, we trained the network from scratch rather than relying on any pre-trained weights. In order to test the predictive power of this network, we applied it to the area of the city of Munich, with the test image being pre-processed with the same GEE-based procedure as described in Section 3.1. The resulting land use map is shown in Figure 6 (top row), and was created by sliding the ResNet-110 on an input Sentinel-2 image with a stride of 10 (leading to an output pixel spacing of 100 m). The overall accuracy (OA), average accuracy (AA) and Kappa coefficient calculated based on a grid of manually annotated control points can be seen in Table 4. It has to be noted that for sake of this evaluation, we combined the classes *Open Forests* (LCCS 20) and *Forest/Cropland Mosaic* (LCCS 25) to a joint *Open Forests* class, and *Natural Herbaceous* (LCCS 30) with *Herbacoeus Croplands* (LCCS 36) and *Natural Herbaceous/Croplands Mosaic* (LCCS 35) to a simple *Herbaceous* class, as it is very difficult for human annotators to distinguish those classes by their only subtle differences.

	ResNet-110	DenseNet
Upsampling type	n/a	Deconvolution
Input size	64 × 64 × 10	256 × 256 × 10
Batch size	16	4
Loss	Categorical cross-entropy	
Optimizer	Nesterov Adam	
Initial LR	0.0005	0.0001
LR schedule	ReduceOnPlateau	

Table 3. Baseline network training configurations.

The second network we used was the fully convolutional DenseNet for semantic segmentation (Jégou et al., 2017), aiming at assigning a class label to every pixel of the input image. Here, we used full-sized Sentinel-2 10-band patches (i.e. 256 × 256 pixels) as input and processed the city of Rome for test purposes. The result is also depicted in Figure 6, and the accuracy metrics, calculated analogue to the Munich case, are described in Table 5.

	OA	Kappa	AA
MODIS	53.0%	0.38	37.5%
Predicted	65.1%	0.51	43.6%

Table 4. Accuracy metrics of the LCCS maps for the Munich scene. The predicted result was achieved by patch classification based on a ResNet-110 CNN. The evaluation is based on manually labeled reference points.

	OA	Kappa	AA
MODIS	56.0%	0.37	33.7%
Predicted	63.3%	0.44	39.4%

Table 5. Accuracy metrics of the LCCS maps for the Rome scene. The predicted result was achieved by semantic segmentation based on a DenseNet CNN. The evaluation is based on manually labeled reference points.

By comparing the resulting maps to the original MODIS-derived LCCS land use map, as well as an image extracted from Google Earth, it can be seen that in both cases the resolution of the map was successfully enhanced so that more details can be retrieved, while an overall agreement between the low-resolution MODIS map and the corresponding predicted high-resolution result is preserved. It has to be highlighed that both test areas are not contained in the *SEN12MS* dataset, so that the results can

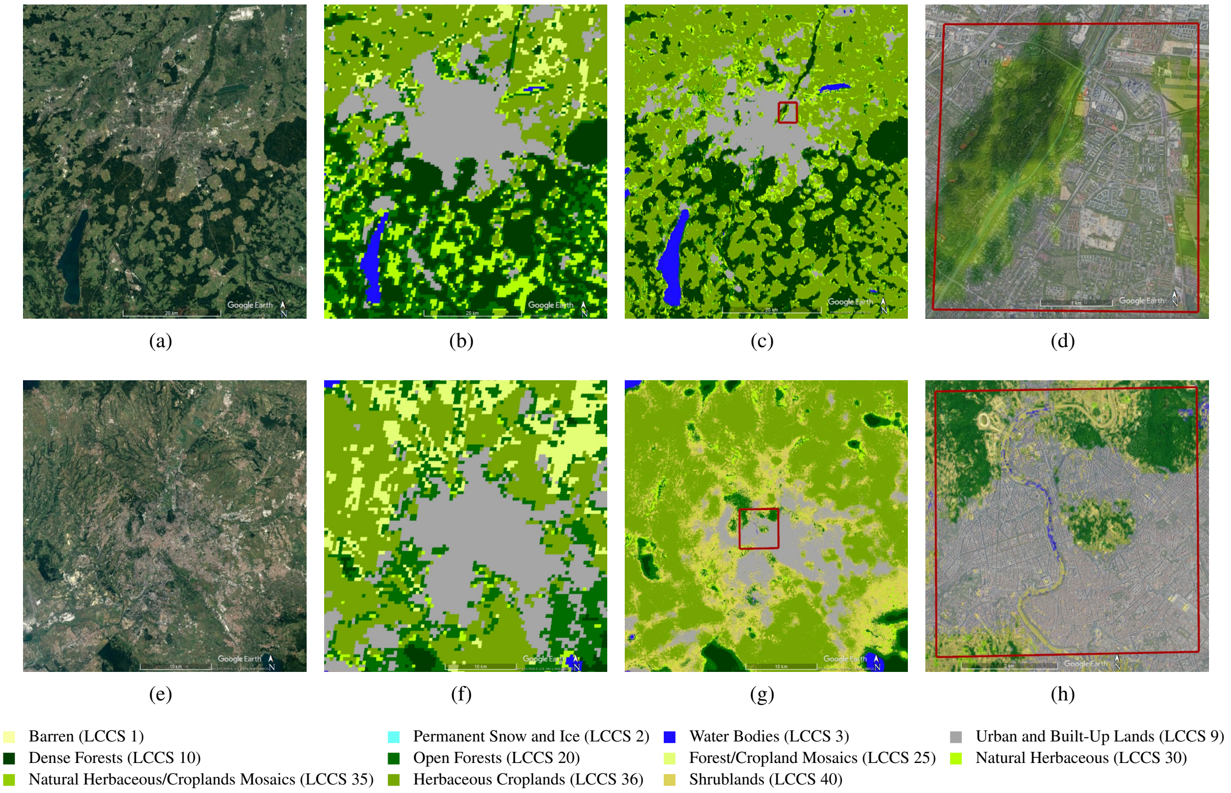
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
Barren (LCCS 1)
Permanent Snow and Ice (LCCS 2)
Water Bodies (LCCS 3)
Urban and Built-Up Lands (LCCS 9)
Dense Forests (LCCS 10)
Open Forests (LCCS 20)
Forest/Cropland Mosaics (LCCS 25)
Natural Herbaceous (LCCS 30)
Natural Herbaceous/Croplands Mosaics (LCCS 35)
Herbaceous Croplands (LCCS 36)
Shrublands (LCCS 40)

Figure 6. Examples from Munich (top) and Rome (bottom) for the predictive power of the dataset: (a) and (e) Optical images extracted from Google Earth, (b) and (f) MODIS-derived LCCS land use map with 500m resolution, (c) predicted LCCS land use map with 100m resolution using classification network, (g) predicted LCCS land use map with 10m resolution using semantic segmentation network, (d) and (h) zoom-in of the red rectangles in the respective predicted LCCS land use maps. While the area in the red rectangle is completely defined as urban in the original MODIS-derived product, more details are visible in the predicted results.

serve as a first indicator of the strong generalization capability provided by the versatility of the dataset. This holds even more so since only a small subset of the dataset (namely the patches of the *summer* season) has been used for training the networks used in the experiments. This impression is also confirmed by the independently evaluated accuracy metrics. We expect that by exploiting the whole dataset, i.e. both sensor modalities, all ROIs and all seasons, powerful classifiers for large-scale mapping applications can be developed. Besides that, it has to be mentioned that training on low-resolution labels of course creates a form of label noise. Using specifically adapted strategies aiming at so-called label super-resolution (Malkin et al., 2019) might be able to provide even better results.

6. DISCUSSION

As can be seen from Table 1, *SEN12MS* is among the five largest datasets when only the sheer number of image patches is considered. However, in the end, *SEN12MS* contains a lot more data and is consequently much bigger than its competitors: First and foremost, its patches are of size 256 × 256 pixels instead of just 28×28 (SAT-4/6) or 120×120 (BigEarthNet) pixels. Besides, the spectral information content is also much higher, as *SEN12MS* contains full multi-spectral Sentinel-2 imagery and, in addition, dual-polarimetric Sentinel-1 SAR data, while most other datasets – besides EuroSAT, SEN1-2, and BigEarthNet – only contain color imagery with 3 or 4 bands. Last, but not least, it has to be mentioned that *SEN12MS* is the most versatile dataset regarding scene distribution, as it covers all regions of the Earth over all meteorological seasons, while most of the other datasets are restricted to fairly small areas (e.g. Brazilian Coffee Scenes or USGS SIRI-WHU), individual countries (e.g. SAT-4/5 or DeepGlobe - Road Extraction), or a single continent (e.g. BigEarthNet).

Wile this versatility is the major strength of *SEN12MS*, it also poses the greatest challenge: In comparison to, e.g., ImageNet, which also provides a lot of versatility and thus tries to model the whole world, the number of samples in *SEN12MS* is still fairly small. As can be seen from the example results provided in Section 5, the dataset nevertheless seems to hold the potential to train powerful, well-generalizing models even under data-scarce training situations. Besides, we believe that further benefit will arise from combining the existing datasets, e.g. BigEarthNet, with *SEN12MS* in the frame of transfer learning to enlarge the amount of data a model can learn patterns from.

7. SUMMARY AND CONCLUSION

With this paper, we publish the *SEN12MS* dataset, which contains 180,662 triplets of Sentinel-1 dual-polarimetric SAR data, Sentinel-2 multi-spectral images, and MODIS-derived land cover maps. With its large patch size, its global scene distribution, and its wealth of versatile remote sensing information, it can be considered to be the largest remote sensing dataset available to date. We hope it will foster the development of well-generalizing machine learning models for a more sophisticated automatic analysis of Sentinel satellite data.

ACKNOWLEDGEMENTS

This work is jointly supported by the Helmholtz Association under the framework of the Young Investigators Group SiPEO (VH-NG-1018), the German Research Foundation (DFG, grant SCHM 3322/1-1), and the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement ERC-2016-StG-714087, Acronym: So2Sat).

REFERENCES

Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., Nemani, R., 2015. DeepSat: a learning framework for satellite imagery. In: *Proc. ACM SIGSPATIAL*. Art. No. 37.

Baumgardner, M.F., Biehl, L.L., Landgrebe, D.A., 2015. 220 band AVIRIS hyperspectral im-age data set: June 12, 1992 Indian Pine Test Site 3. https://purr.purdue.edu/publications/1947/1.

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raska, R., 2018. DeepGlobe 2018: A challenge to parse the earth through satellite images. In: *Proc. CVPR Workshops*, 172–181.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database. In: *Proc. CVPR*, 248–255.

Di Gregorio, A., 2005. *Land cover classification system: classification concepts and user manual: LCCS*. Food & Agriculture Org.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P. et al., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.*, 120, 25–36.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.*, 202, 18–27.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proc. CVPR*, 770–778.

Helber, P., Bischke, B., Dengel, A., Borth, D., 2018. Introducing EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification. In: *Proc. IGARSS*, 204–207.

Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., Bengio,Y., 2017. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In: *Proc. CVPR Workshops*, 11–19.

Loveland, T., Belward, A., 1997. The international geosphere biosphere programme data and information system global land cover data set (DISCover). *Acta Astronautica*, 41(4-10), 681–689.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017.Can semantic labeling methods generalize to any city? The Inria aerial image labeling benchmark. In: *Proc. IGARSS*, 3226–3229.

Malkin, K., Robinson, C., Hou, L., Soobitsky, R., Samaras, D., Saltz, J., Joppa, L., Jojic, N., 2019. Label super-resolution networks. In: *Proc. ICLR*, 1–9.

Mohajerani, S., Saeedi, P., 2019. Cloud-Net: An end-to-end cloud detection algorithm for Landsat 8 imagery. *arXiv:1901.10077*.

Nogueira, K., Dos Santos, J. A., Fornazari, T., Silva, T. S. F., Morellato, L. P., Torres, R., 2016. Towards vegetation species discrimination by using data-driven descriptors. In: *Proc. PRRS*.

Penatti, O. A. B., Nogueira, K., dos Santos, J. A., 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In: *Proc. CVPR Workshops*, 44–51.

Rieke, C., 2019. Awesome satellite imagery datasets. `https://github.com/chrieke/awesome-satellite-imagery-datasets`. [Online; accessed 15-March-2019].

Schmitt, M., Hughes, L.H., Zhu, X.X., 2018. The SEN1-2 dataset for deep learning in SAR-optical data fusion. In: *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-1, 141–146.

Schmitt, M., Hughes, L.H., Qiu, C., Zhu, X.X., 2019. Aggregating cloud-free Sentinel-2 images with Google Earth Engine. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.

Schubert, A., Small, D., Miranda, N., Geudtner, D., Meier, E., 2015. Sentinel-1A product geolocation accuracy: Commissioning phase results. *Remote Sensing*, 7(7), 9431–9449.

Sulla-Menashe, D. and Friedl, M. A., 2018. User guide to Collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product. *USGS: Reston, VA, USA*.

Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P. and Friedl, M. A., 2019. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 land cover product. *Remote Sens. Environ.*, 222, 183 – 194.

Sumbul, G., Charfuelan, M., Demir, B. and Markl, V., 2019. BigEarthNet: A large-scale benchmark archive for remote sensing image understanding. *arXiv:1902.06148*.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M. et al., 2012. GMES Sentinel-1 mission. *Remote Sens. Environ.*,120, 9–24.

Warmerdam, F., 2008. *The Geospatial Data Abstraction Library*. Springer, Berlin, 87–104.

Xia, G., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2018. Dota: A large-scale datasetfor object detection in aerial images. In: *Proc. CVPR*, 3974–3983.

Yang, Y., Newsam, S., 2010. Bag-of-visual-words and spatial extensions for land-use classification. In: *Proc. ACM SIGSPATIAL*, 270–279.

Yokoya, N., Ghamisi, P., Xia, J., Sukhanov, S., Heremans, R., Tankoyeu, I., Bechtel, B., Saux, B. L., Moser, G., Tuia, D., 2018. Open data for global multimodal land use classification: Outcome of the 2017 IEEE GRSS data fusion contest. *IEEE J. Sel. Topics Appl. Earth Observ.,* 11(5),1363–1377.

Zhao, B., Zhong, Y., Xia, G., Zhang, L., 2016. Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.,* 54(4), 2108–2123.

Zhong, Y., Zhu, Q., Zhang, L., 2015. Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.,* 53(11), 6207–6222.

Zhu, X.X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.,* 5(4), 8–36.