MULTISENGE : A MULTIMODAL AND MULTITEMPORAL BENCHMARK DATASET FOR LAND USE/LAND COVER REMOTE SENSING APPLICATIONS

Romain Wenger*, Anne Puissant*, Jonathan Weberb, Lhassane Idoumgharb and Germain Forestiera

*aLIVE UMR 7362 CNRS, University of Strasbourg, F-67000 Strasbourg, France (romain.wenger; anne.puissant)@live-cnrs.unistra.fr
bIRIMAS UR 7499, University of Haute-Alsace, F-68100 Mulhouse, France (jonathan.weber; lhassane.idoumghar; germain.forestier)@uha.fr

KEY WORDS: Benchmark dataset, Deep Learning, Semantic segmentation, Scene classification, Sentinel archive, Landuse/cover

ABSTRACT:

This paper presents MultiSenGE that is a new large scale multimodal and multitemporal benchmark dataset covering one of the biggest administrative region located in the Eastern part of France. MultiSenGE contains 8,157 patches of 256 × 256 pixels for the Sentinel-2 L2A, Sentinel-1 GRD images in VV-VH polarization and a Regional large scale Land Use/Land Cover (LULC) topographic reference database. With MultiSenGE, we contribute to the recent developments towards shared data use and machine learning methods in the field of environmental science. The purpose of this dataset is to propose relevant and easy-access dataset to explore deep learning methods. We use MultiSenGE to evaluate the performance for urban areas using well-known deep learning techniques. These results serve as a baseline for future research on remote sensing applications using the multi-temporal and multimodal aspects of MultiSenGE. With all patches georeferenced at a 10 meters spatial resolution covering the whole Grand-Est Region, MultiSenGE provides an opportunity for environmental benchmark dataset will help to advance data-driven techniques for land use/land cover remote sensing applications.

1. INTRODUCTION

The constant evolution of earth observation missions has allowed the acquisition of a large amount of satellite data. They have considerably changed the way humanity manages its territory. One of the major examples is the Copernicus program developed by the European Space Agency (ESA) which consists in the deployment of several constellations of satellites to monitor the Earth Surface. This is the case with the Sentinel missions, which currently consists of six missions designed as a two-satellite constellation. Sentinel-1 and Sentinel-2 provide open-access and freely Synthetic Aperture Radar (SAR) and multispectral imagery with a very short revisit time, respectively 5 days for Sentinel-2A/2B and between 5 and 10 days depending on the location on Earth for Sentinel-1A/1B.

The high revisit period of these two sensors allows the use of time series to study the dynamics and evolution of processes and objects of interest. Satellite Image Time Series (SITS) have already been used for the analysis of agricultural areas (Bégüe et al., 2018, Bellón et al., 2017, Kussul et al., 2017, Rußwurm et al., 2020), forest areas (Wulder et al., 2012, Pickell et al., 2016) or classification of Land Use/Land Cover (LULC) (Inglada et al., 2017). In addition, the joint use of SAR and optical data continues to increase the interest of researchers, especially in the case of LULC mapping (Ienco et al., 2019, Steinhausen et al., 2018). This combination has already shown its performance in many other works such as the detection of natural areas (Dusseux et al., 2014, Mngadi et al., 2021), the detection of changes (Gao et al., 2017) or the mapping of urban areas (Iannelli and Gamba, 2018).

To deal with this increasing amount of data, new techniques based on neural networks have been developed and offer promising results in the classification of LULC from multiple data sources (Ma et al., 2019). Multimodal and multitemporal datasets are currently quite rare and remain for the most part specific to applications (Object detection, Scene classification, Semantic segmentation or Instance segmentation) and it is important to have a variety of data sets for the application of deep learning models. To our knowledge, only two datasets use Sentinel-1/Sentinel-2 pairs for scene classification or semantic segmentation, BigEarthNet (Sumbul et al., 2021) and SEN12MS (Schmitt et al., 2019). BigEarthNet proposes 590,326 pairs of single-time annotated images with the Corine Land Cover reference data over 10 countries of Europe. SEN12MS offers 180,662 Sentinel-1 and 2 triplets and MODIS Land Cover over several regions in the world and for spring, summer and winter seasons to perform semantic segmentation.

In order to support the lack of multimodal and multitemporal datasets, we decided to produce MultiSenGE, a benchmark dataset covering the Grand-Est region in France (57,433 km² which represents 10.6% of the French territory). The objective is to focus the benchmark dataset on semantic segmentation and classification. This dataset offers Sentinel-1, Sentinel-2 and LULC triplets, a land cover data recently available with an open-source licence over this territory. Compared to other existing datasets, MultiSenGE allows to classify urban surfaces into 5 LULC classes, against only 1 for SEN12MS using MODIS Land Cover as reference data and 11 for BigEarthNet using CORINE Land Cover as reference data. We use OCSGE2-GEORANDEST which have the advantage to own a Minimum Mapping Unit (MMU) less than 50m², which gives it a higher geometric accuracy than existing LULC products and is close and consistent with the spatial resolution of Sentinel imagery (10m).

First is presented the satellite and LULC data on the study area used to build the dataset. Then the methodology to process the reference data and to create triplets of patches is presented. Fi-
nally, baseline results performed on the beta version of MultiS- 
tenGE only based on urban thematic classes are described before conclusion and perspectives.

2. REFERENCE DATA AND SATELLITE IMAGERY

2.1 Study sites and reference data

The dataset covers a large territory (57,433 km²) in eastern France and corresponds to an administrative French district (Figure 1) extended from Alsace in the East to the Ar- dennes and Marne in the West. This area have been chosen due to the availability of a new, accurate and up-to-date vector LULC database named OCSGE2-GEORANDEST (www.georandest.fr). This open-access vector database was built by visual interpretation of aerial photographs for 2019/2020. It is organized into four levels of nomenclature where the first level categorizes land cover into four classes (1) artificial surfaces, (2) agricultural areas, (3) forest areas, and (4) water surfaces. At the most accurate level (1:10,000), 53 LULC classes map the region and the size of the smallest elements is 50m². In order to obtain a generic reference data with 14 classes and have class consistency at 10m spatial resolution, several preprocessing steps are performed on the original topographic vector database (see 3.1). In the OCSGE2 layer, all the roads have the same degree of importance. Some of them are too small to be distinguished at 10 m spatial resolution. Then, in order to produce an adapted road network label, a second database (BDTOPO-IGN), produced by IGN describing lines in vector format, with the degree of importance, is pre-processed.

2.2 Sentinel-2

Sentinel-2 offers 13 spectral bands at 10m and 20m spatial resolution. The Satellite Image Time Series (SITS) for 2020 (14 tiles - Figure 1) are downloaded from the Theia land services and datacenter (https://www.theia-land.fr/), in L2A format, corrected from atmospheric effects with a cloud mask (surface reflectance product). Only free cloud cover images (with cloud cover strictly less than 10%) are automatically downloaded from the Theia land services (https://www.theia-land.fr/), in L2A format, corrected from atmospheric effects with a cloud mask (surface reflectance product). Only free cloud cover images (with cloud cover strictly less than 10%) are automatically downloaded from the Theia land services (https://www.theia-land.fr/), in L2A format, corrected from atmospheric effects with a cloud mask (surface reflectance product). Only free cloud cover images (with cloud cover strictly less than 10%) are automatically downloaded from the Theia land services (https://www.theia-land.fr/), in L2A format, corrected from atmospheric effects with a cloud mask (surface reflectance product). Only free cloud cover images (with cloud cover strictly less than 10%) are automatically downloaded from the Theia land services (https://www.theia-land.fr/), in L2A format, corrected from atmospheric effects with a cloud mask (surface reflectance product). Only free cloud cover images (with cloud cover strictly less than 10%) are automatically downloaded from the Theia land services (https://www.theia-land.fr/).

2.3 Sentinel-1

Sentinel-1 is equipped with C-band SAR sensors which allows the acquisition of imagery day and night without weather disturbances compared to optical imagery. It provides data in dual polarization with two product types: Grand Range Detected (GRD) and Single Look Complex (SLC) (Filipponi, 2019). GRD products, used to construct this dataset, consist on SAR data that have been multi-looked and projected to ground range using Earth ellipsoid model.

The SAR images available in ascending and descending orbits for 2020 were downloaded and pre-processed using the S1-Tiling (Koleck and Centre National des Etudes Spatiales (CNES), 2021) processing chain developed by CNES (Centre National d’Etudes Spatiales). This processing chain can be divided into four points: (1) automatic downloading of Sentinel-1 data thanks to the EODAG library which offers the possibility to request different servers to always have data available on the studied area, (2) slicing of the SAR data according to the Sentinel-2 tiling, (3) orthorectification of the newly sliced SAR scenes and (4) application of a multi-temporal filter to reduce the speckle and preserve the spatial information.

3. MULTISENGE PRODUCTION

3.1 Reference data processing

In order to obtain LULC data at 10 meters spatial resolution, five pre-processing steps are applied on the reference data (Figure 2). The five steps consist in (1) resampling reference labeled data, (2) removing the smallest polygons thanks to the connected component labeling method applied on each selected class, (3) filling the holes resulting from this method by nearest neighbor, (4) applying a mathematical morphology of closure to smooth the outlines of each class on the final data and (5) adding roads from the second database.

First, to obtain a number of labeled classes adapted to the spatial resolution, we kept the classes at level 3 of the nomenclature with some semantic reclassification to reduce the complexity of the typology, especially for semantic segmentation applications. Finally, the final LULC typology includes 14 classes with 5 classes for urban areas and 9 classes for natural surfaces (Table 1).

The second step is applied to reduce the spatial complexity of the reference data which can contain small polygons (50m²)
Table 1. Class name and pixel value in the final reference data in a raster format.

| Original level 1 typology | New level 3 typology |
|---------------------------|----------------------|
| Urban Areas (1)           | Dense Built-Up (1)   |
|                           | Sparse Built-Up (2)  |
|                           | Specialized Built-Up Areas (3) |
|                           | Specialized but Vegetative Areas (4) |
|                           | Large Scale Networks (5) |
| Agricultural areas (2)    | Arable Lands (6)     |
|                           | Vineyards (7)        |
|                           | Orchards (8)         |
|                           | Grasslands (9)       |
|                           | Groces, Hedges (10)  |
| Forests and semi-natural areas (3) | Forests (11) |
| Wetlands (4)              | Wetlands (13)        |
| Water Surfaces (5)        | Water Surfaces (14)  |

3.2 Triplet data preparation

The Sentinel-1 SAR SITS, Sentinel-2 optical SITS and the reference data are cut into 256 x 256 pixel patches. The VV and VH bands from the pre-processed Sentinel-1 images are stacked for each patch. For Sentinel-2, the bands at 10 meters are kept and the bands at 20 meters are resampled to 10 meters spatial resolution using cubic interpolation to have homogeneity in the final data. The 10 bands are then stacked for each patch. Finally, the simplified LULC reference data is also cut following its typology (Figure 4) will also be available for download.

MultiSenGE contains four folders with (1) the simplified reference data patches, (2) the Sentinel-2 patches, (3) the Sentinel-1 patches and (4) the GeoJSON files. The files contained in the folders are identified according to the following nomenclatures:

- `ground_reference` : 
  ```
  {tile},GR,{x-pixel-coordinate},{y-pixel-coordinate}.tif
  ```
- `s2` : 
  ```
  {tile},{date},S2,{x-pixel-coordinate},{y-pixel-coordinate}.tif
  ```
- `s1` : 
  ```
  {tile},{date},S1,{x-pixel-coordinate},{y-pixel-coordinate}.tif
  ```
- `labels` : 
  ```
  {tile},{x-pixel-coordinate},{y-pixel-coordinate}.json
  ```

Each triplet is described by labels to perform both scene classification and semantic segmentation (Figure 5). In addition to the classes, this GeoJSON file also contains the names of all associated Sentinel-2 and Sentinel-1 patches as well as the specific projection for each patch. The projection of each patch is in UTM/WGS84 format inherited from its original tile. For the current version, the dataset contains 8,157 non-overlapping triplets along with the GeoJSON file containing all the classes present in the patch. The mosaic level 3 reference data product and its typology (Figure 4) will also be available for download.
where tile is the Sentinel-2 tile number, x-pixel-coordinate and y-pixel-coordinate are the coordinates of the patch in the tile and date is the date of acquisition of the patch image as the time series is extracted from each sensor. GR means Ground Reference and correspond to the ground reference patches, S2 Sentinel-2 and correspond to the Sentinel-2 patches and S1 Sentinel-1 for the Sentinel-1 patches. Users will find 72,033 multi-temporal patches for Sentinel-2 and 1,012,227 multi-temporal patches for Sentinel-1. To facilitate the reading and extraction of dates from the dataset, tools have been developed and are available on a code hosting service.

4. BASELINE RESULTS

First baseline results are performed on urban areas using only single-time patches and the five urban classes of the dataset (Table 1, classes 1 to 5). First experiments focused on the Moselle department (T31UGQ). We then selected a subset on the tile for training and validation steps and a different spatially tile for testing zone (Saraiva et al., 2020). Two multiclass U-Net networks with VGG-16 as backbone are tested split subset for testing zone (Saraiva et al., 2020). Two multiclass networks for baseline results over Metz.

The statistical results show better scores for the U-Net-IRRG method compared to the U-Net-Index method with weighted $F_1$ Score of 0.7364 and 0.7214 respectively (Table 3). Moreover, four of the six classes studied have a better $F_1$ Score for U-Net-IRRG than U-Net-Index (Table 4). These statistical results are confirmed by the visual results presented in Figure 6. We notice a better homogeneity in the classification of all the classes on the test area for the first method while the second method offers a much more fragmented classification with an underestimation of several classes.

These baseline models are the first ones applied only on a single tile (T31UGQ). Others semantic segmentation deep learning methods using the multimodal and multitemporal dataset are a part of an ongoing PhD research.

5. CONCLUSION

This paper presents MultiSenGE, a new large scale multimodal and multitemporal benchmark dataset covering a large area.
in the eastern of France. It contains 8,157 triplets of 256 x 256 pixels of Sentinel-1 dual-polarimetric SAR data, Sentinel-2 multispectral level 2A images and a LULC reference database built to obtain a typopoly adapted to the 10m spatial resolution for year 2020. Moreover, the proposed 14 classes are consistent to map french or european national-scale landscape. Triplets data offer users the possibility to perform semantic segmentation and scene classification on multitemporal and multimodal data with large input patches. The baseline results have shown its ability to offer encouraging first results on urban areas classification with patches only with the single-time optical imagery. Others tests are ongoing and should improve these first results by using multitemporal and multimodal imagery offered by this dataset as well as different deep learning techniques. This work is supported by French Research Agency (ANR-17-CE23-0015) and is part of a PhD project.

Table 4. Classes $F_1$-Score for baseline results over Metz.

| Classes | $F_1$-Score |
|---------|-------------|
| Class 1 | 0.5199      |
| Class 2 | 0.6565      |
| Class 3 | 0.5090      |
| Class 4 | 0.3141      |
| Class 5 | 0.5862      |
| Class 6 | 0.8549      |

**REFERENCES**

Audebert, N., Le Saux, B., Lefèvre, S., 2018. Beyond RGB: Very High Resolution Urban Remote Sensing With Multimodal Deep Networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140, 20-32. https://hal.archives-ouvertes.fr/hal-01636145.

Bellón, B., Bégué, A., Lo Seen, D., De Almeida, C. A., Simões, M., 2017. A Remote Sensing Approach for Regional-Scale Mapping of Agricultural Land-Use Systems Based on NDVI Time Series. *Remote Sensing*, 9(6).

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., De Avelleyra, D., P. D. Ferraz, R., Lebourgeois, V., Lelong, C., Simões, M., R. Vérón, S., 2018. Remote Sensing and Cropping Practices: A Review. *Remote Sensing*, 10(1).

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition, Ieee, 248–255.

Dusseux, P., Corpetti, T., Hubert-Moy, L., Corgne, S., 2014. Combined Use of Multi-Temporal Optical and Radar Satellite Images for Grassland Monitoring. *Remote Sensing*, 6(7), 6163–6182.

Filipponi, F., 2019. Sentinel-1 GRD Preprocessing Workflow. *Proceedings*, 18(1). https://www.mdpi.com/2504-3900/18/1/11.

Gao, Q., Zribi, M., Escorihuela, M. J., Baghdadi, N., 2017. Synergistic Use of Sentinel-1 and Sentinel-2 Data for Soil Moisture Mapping at 100 m Resolution. *Sensors*, 17(9).

Hagolle, O., 2021. theia_download. https://github.com/olivierhagolle/theia_download (21 February 2020).

Haralick, R. M., Dinstein, I., Shanmugam, K., 1973. Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6), 610–621.

Iannelli, G. C., Gamba, P., 2018. Jointly Exploiting Sentinel-1 and Sentinel-2 for Urban Mapping. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 8209-8212.

Ienco, D., Interdonato, R., Gaetano, R., Ho Tong Minh, D., 2019. Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158, 11-22.

**ACKNOWLEDGEMENTS**

We thanks the Spatial Data Infrastructure GeoGrandEst provided the reference data used in this study and the Theia Services and Data Infrastructure for the Sentinel-2A imagery. We also would like to thanks Peps Platform for the Sentinel-1 data with large input patches. The baseline results have shown its ability to offer encouraging first results on urban areas classification with patches only with the single-time optical imagery. Others tests are ongoing and should improve these first results by using multitemporal and multimodal imagery offered by this dataset as well as different deep learning techniques. This work is supported by French Research Agency (ANR-17-CE23-0015) and is part of a PhD project.

Figure 6. Visual results for two subsets over Metz. (a) is the Sentinel-2 RGB, (b) is the ground reference in a raster format, (c) U-Net-IRRG prediction and (d) U-Net-index prediction.
