



OPEN

CatLC: Catalonia Multiresolution Land Cover Dataset

DATA DESCRIPTOR

Carlos García^{1,2}, Oscar Mora¹✉, Fernando Pérez-Aragüés¹ & Jordi Vitrià²

The availability of large annotated image datasets represented one of the tipping points in the progress of object recognition in the realm of natural images, but other important visual spaces are still lacking this asset. In the case of remote sensing, only a few richly annotated datasets covering small areas are available. In this paper, we present the Catalonia Multiresolution Land Cover Dataset (CatLC), a remote sensing dataset corresponding to a mid-size geographical area which has been carefully annotated with a large variety of land cover classes. The dataset includes pre-processed images from the Cartographic and Geological Institute of Catalonia (ICGC) (<https://www.icgc.cat/en/Downloads>) and the European Space Agency (ESA) (<https://scihub.copernicus.eu>) catalogs, captured from both aircraft and satellites. Detailed topographic layers inferred from other sensors are also included. CatLC is a multiresolution, multimodal, multitemporal dataset, that can be readily used by the machine learning community to explore new classification techniques for land cover mapping in different scenarios such as area estimation in forest inventories, hydrologic studies involving microclimatic variables or geologic hazards identification and assessment. Moreover, remote sensing data present some specific characteristics that are not shared by natural images and that have been seldom explored. In this vein, CatLC dataset aims to engage with computer vision experts interested in remote sensing and also stimulate new research and development in the field of machine learning.

Background & Summary

Pixel-wise classification of remote sensing images is a challenging task that often requires fieldwork and manual annotation because of the importance of its role in several critical applications. Therefore, mapping agencies and organizations are on the quest to explore how to minimize these strenuous and time-consuming manual tasks by using computer-assisted processes. To do so, current automatic land cover segmentation techniques have benefited from better raw data sources, but they still require improvement in terms of accuracy and also better integration strategies with humans in the loop.

Land cover mapping is among the primary use cases of airborne and satellite images and the proposed focus of this work. Land cover maps are used in different applications such as forest inventory and management, hydrology, crop management or geologic risk identification and assessment. Therefore, accurate and updated knowledge about land dynamics is essential for territory management with different purposes and in multiple fields but, nowadays, high-resolution land segmentation is still done mainly by employing photointerpretation techniques, entailing high costs in terms of time and human resources.

The transformation towards a computer-assisted solution faces a critical point: the scarcity of high-quality datasets, being the labeling process one of the main causes of this situation. Labeling natural image datasets, such as ImageNet¹ and PASCAL VOC², does not pose an interpretation problem as their classes are well defined, distinctive, and can be easily understood by any human annotator. However, labeling remote sensing images correctly might need expert knowledge and requires access to different sources of raw data. For example, differentiating between deciduous or evergreen forest is not an easy task even for the expert.

When data are scarce, the best strategy when developing classification models is to adapt models that have been developed in similar fields with plenty of data. The case of remote sensing is not an exception, and the use of traditional natural image segmentation architectures is the paradigm in the field. There are some reasons to think that these models are not optimal for remote sensing images because of their inadequate inductive biases, but this hypothesis can only be validated by having access to large datasets of carefully labeled remote sensing data. It is necessary to tackle issues such as the restricted translational invariance of these images or the variable resolution of their bands.

¹Institut Cartogràfic i Geològic de Catalunya, Barcelona, 08038, Spain. ²Universitat de Barcelona, Departament de Matemàtiques i Informàtica, Barcelona, 08007, Spain. ✉e-mail: oscar.mora@icgc.cat



Fig. 1 Continuous area using different layers of the dataset together with the ground truth labels.



Fig. 2 Location of the area of interest, Catalonia (Spain).

Additionally, in the near future, full automation of high-resolution cartographic tasks such as land mapping will be the norm, and better strategies to develop powerful deep learning models with a human in the loop are necessary. Traditional pixel-wise deep learning segmentation techniques must also be adapted to this end³.

In this paper, we present the Catalonia Multiresolution Land Cover Dataset (CatLC). This dataset (see Fig. 1) comprises a large variety of images: RGB and infrared orthophotos from airborne sensors at high resolution, radar imagery from Sentinel-1 satellites, multispectral data from Sentinel-2 satellites and compositions of topographic maps—all those accompanied by a land cover map labeled by experts in photointerpretation. Using different combinations of images from the dataset, we offer a benchmark that could serve as a starting point to explore different artificial intelligence techniques for remote sensing segmentation purposes. CatLC dataset aims to engage with computer vision experts interested in remote sensing and stimulate research and development.

Methods

In this section, the CatLC dataset is presented in detail. It includes a set of images, obtained by airborne and satellite sensors, from the catalogs of the Cartographic and Geological Institute of Catalonia (ICGC) and the European Space Agency (ESA). Labels correspond to the current ICGC's land cover map.

The CatLC dataset covers the entire territory of Catalonia (Spain) (see Fig. 2), approximately 32000 km², providing a high quality source of information for the application of Artificial Intelligence (AI) and Deep Learning (DL) techniques, both regarding the variety of the information and their extension.

All the images were acquired during 2018 at different spectral bands and spatial resolutions. They are provided in GeoTiff raster format and share a common georeferencing system projection (WGS84 UTM31N Reference System). They cover the same geographic extension, given by the following bounding box: UTM X West: 240000, UTM X East: 540000, UTM Y North: 4780000, and UTM Y South: 4480000.

The different data layers, with spatial resolutions varying between 1 and 10 meters depending on the product and sensor used, are presented in detail in the following subsections. In such subsections we illustrate the images provided by the dataset (land cover, orthophoto, Sentinel-1, Sentinel-2 and topographic maps). Figures 5, 6, 7, 8, 9, 11, 12 and 13 show different images on three geographical areas in Catalonia. We have summarized all available data in Table 1.

Land cover map. The 2018 land cover map presented here has 41 different classes (see Fig. 3), including different agricultural areas, forest areas, urban areas and water bodies. Photointerpreters from ICGC followed a standardized procedure during its generation process. The minimum area for labeling an element was 500

Agricultural area			Urban area			Water bodies		
01	Herbaceous crops	14	Meadows and grasslands	28	Sports and leisure areas			
02	Orchard, plant nurseries...	15	Shore forest	29	Mining or landfills			
03	Vineyards	16	Bare forest soil	30	Areas in transformation			
04	Olive groves	17	Burned areas	31	Road network			
05	Other woody crops	18	Rocky	32	Urban bare ground			
06	Crops in transformation	19	Beaches	33	Airport areas			
Forest area			20	Wetlands	34	Railway network		
07	Dense coniferous forests	Urban area			35	Port areas		
08	Dense deciduous forests	21	Urban area	Water bodies				
09	Dense forests of sclerophylls	22	"Eixample"	36	Reservoir			
10	Scrub	23	Lax Urban Areas	37	Lakes and lagoons			
11	Clear coniferous forests	24	Isolated buildings	38	Watercourses			
12	Clear deciduous forests	25	Isolated residential areas	39	Rafts			
13	Clear forests of sclerophylls	26	Green areas	40	Artificial channels			
		27	Industrial or commercial	41	Sea			

Fig. 3 CatLC dataset with 41 classes and legend.

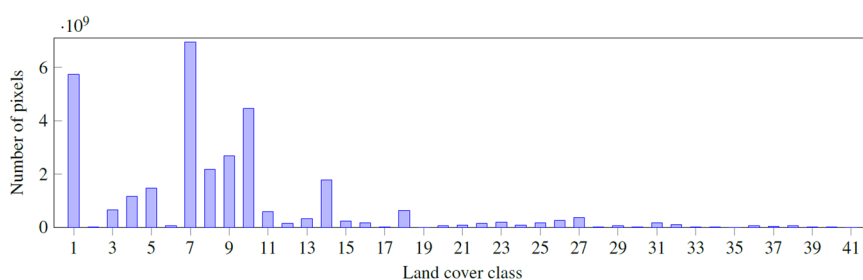


Fig. 4 Class distribution on CatLC dataset.

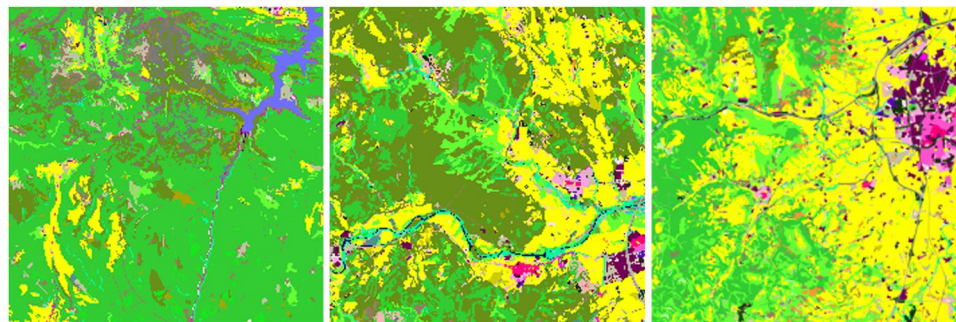


Fig. 5 The distribution of the land covers within the mapped territory is heterogeneous. Some covers as herbaceous crops or dense coniferous forests are much more common than airport areas or water bodies. In Fig. 4, we can see the histogram for the complete dataset.

squared meters and the minimum length for linear features such as roads, rivers, railroad tracks, etc. was between 8 and 10 meters (<https://datacloud.ide.cat/espificacions/cobertes-sol-v1r0-esp-01ca-20160919.pdf>).

Supervision has been performed on a sample of 811 points throughout this territory, resulting in a thematic accuracy of 81%. The final 41 labels (see Fig. 3) presented in this publication are delivered at spatial resolution of 1 m.

The distribution of the land covers within the mapped territory is heterogeneous. Some covers as herbaceous crops or dense coniferous forests are much more common than airport areas or water bodies. In Fig. 4 we can see the histogram for the complete dataset (see also Fig. 5).

Orthophoto. An orthophoto is a cartographic document consisting of a vertical aerial image that has been rectified in such a way as to maintain a uniform scale over the entire image surface. It consists of a geometric representation at a given scale of the Earth's surface.

Original images were taken with a resolution of 25 centimeters, but because the land cover map has a resolution of 1 meter, we have decided to rescale the orthophoto raster layer also to 1 meter.



Fig. 6 Orthophoto RGB samples.

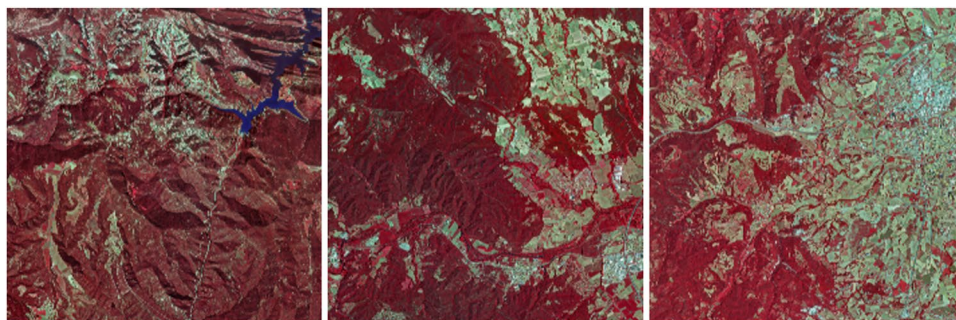


Fig. 7 Orthophoto (Infrared,R,G) samples.

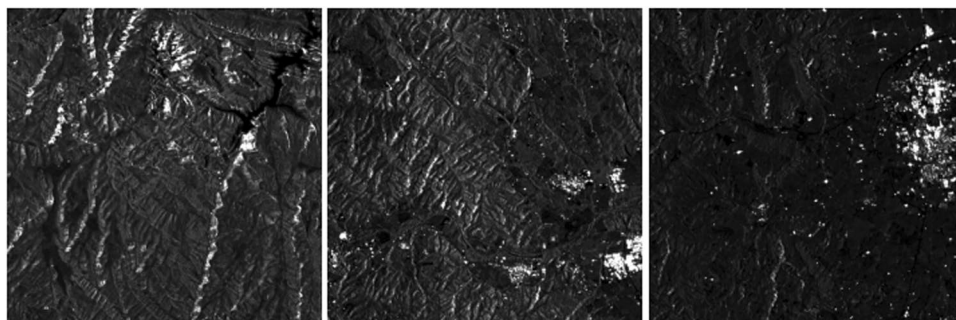


Fig. 8 Sentinel-1 (average image during 2018) samples.



Fig. 9 Sentinel-2 RGB April (a–c) and August (d–f) 2018 samples.

This layer comprises four distinct bands, each providing information from different zones of the electromagnetic spectrum. Three of them belong to the visible area of the spectrum (RGB) (see Fig. 6) and one of them to the infrared area (see Fig. 7). A continuous image is generated based on several thousands of independent photoshoots processed with a combination of commercial software (Trimble/Inpho) and in-house developments. On this cartographic document, digital retouching tasks have been carried out to minimize artifacts that may have originated during the acquisition and processing of the images. The applicable specification can be found in⁴.

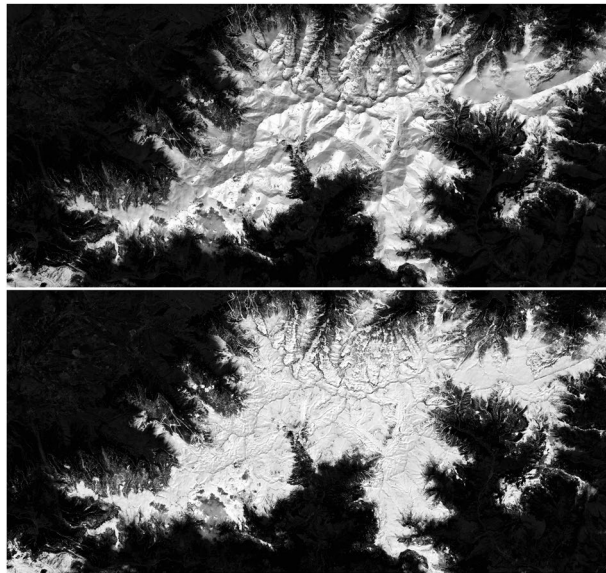


Fig. 10 Sentinel-2 process with atmospheric and topographic corrections. Original (up) and corrected image (down).

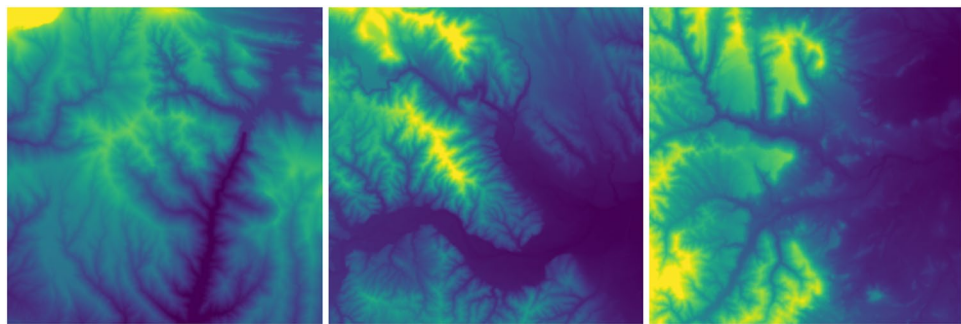


Fig. 11 Digital Elevation Model (DEM) samples.

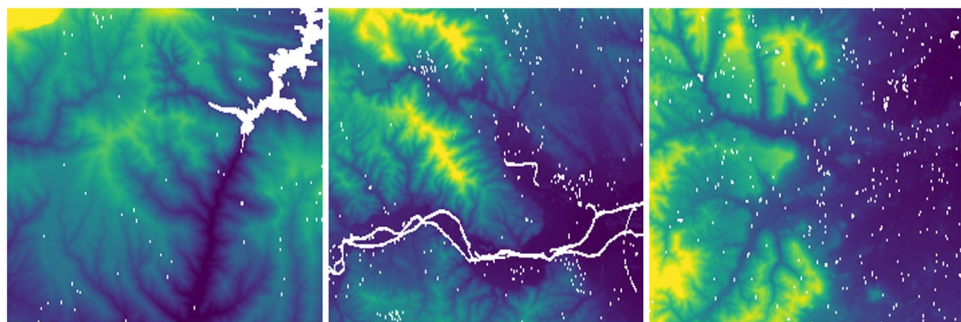


Fig. 12 Digital Surface Model (DSM) samples.

Sentinel-1. The Sentinel-1 dataset has been generated from Synthetic Aperture Radar (SAR) images in GRD (Ground Range Detected) mode from the year 2018 at 10-meter spatial resolution. The Sentinel-1 satellite constellation is made up of two twin satellites, A and B, from the European Space Agency (ESA). These satellites emit a microwave signal (frequency 5.405 GHz) and subsequently receive the echo of the reflection on the ground surface. Therefore, Sentinel-1 images contain information on the reflectivity of the terrain that depending on its type (urban, vegetation, crops, water, etc.) will have different intensities, thus providing valuable information for land cover classification. For this purpose, 12 acquisitions have been chosen, one for each month of the year, covering the entire territory of Catalonia. Full coverage has been achieved by combining two orbits in

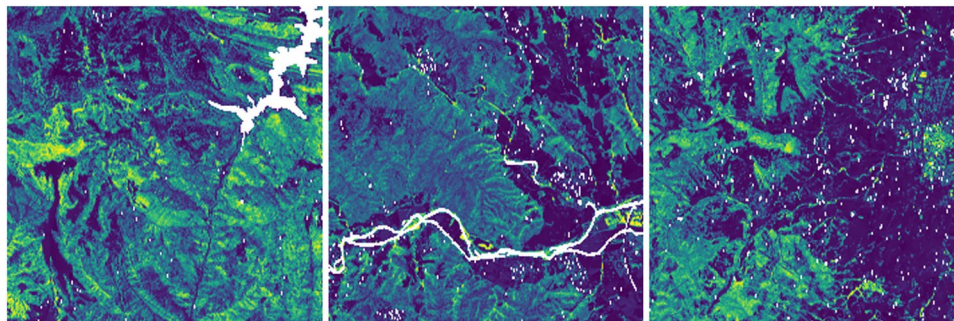


Fig. 13 Canopy Height Model (CHM) samples.

Data	Independent Bands	Resolution (m)
Orthophoto RGB	4	1
Sentinel-1	12	10
Sentinel-2	20	10
DEM	4	5
DSM	4	1
CHM	4	1
Land cover map	—	1

Table 1. Summary of CatLC dataset. 34040 images at 960×960 pixels per source of 1 m data.

Data	Type	Dimensions (pixels)	Pixel Resolution	Size
Orthophoto RGB	GeoTiff	300.000×300.000	1 m	252,3 GB
Orthophoto IR	GeoTiff	300.000×300.000	1 m	84,1 GB
Sentinel-1 January	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-1 February	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-1 March	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-1 April	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-1 May	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-1 June	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-1 July	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-1 August	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-1 September	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-1 October	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-1 November	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-1 December	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-1 Mean 2018	GeoTiff	30.000×30.000	10 m	3,4 GB
Sentinel-2 April	GeoTiff	30.000×30.000	10 m	16,8 GB
Sentinel-2 August	GeoTiff	30.000×30.000	10 m	16,8 GB
Topography - Aspect	GeoTiff	60.000×60.000	5 m	13,4 GB
Topography - CHM	GeoTiff	300.000×300.000	1 m	336,4 GB
Topography - DEM	GeoTiff	60.000×60.000	5 m	13,4 GB
Topography - DSM	GeoTiff	300.000×300.000	1 m	336,4 GB
Topography - Slope	GeoTiff	60.000×60.000	5 m	13,4 GB
Land Cover Map	GeoTiff	300.000×300.000	1 m	84,1 GB

Table 2. Detailed description of the different files in the dataset.

ascending mode (orbits 30 and 132) and VV (Vertical-Vertical) polarizations in similar dates. The descending orbit and VH (Vertical-Horizontal) polarization have not been included in the present dataset because the information is mostly redundant. However, its use can be explored in case it provides improvements in segmentation. Additionally, an average image of the year 2018 has been generated with improved radiometry (multitemporal speckle reduction) by combining all 12 monthly images into one (see Fig. 8). Consequently, the average image cannot provide information on temporary changes during 2018 but does provide a lower noise-level image.

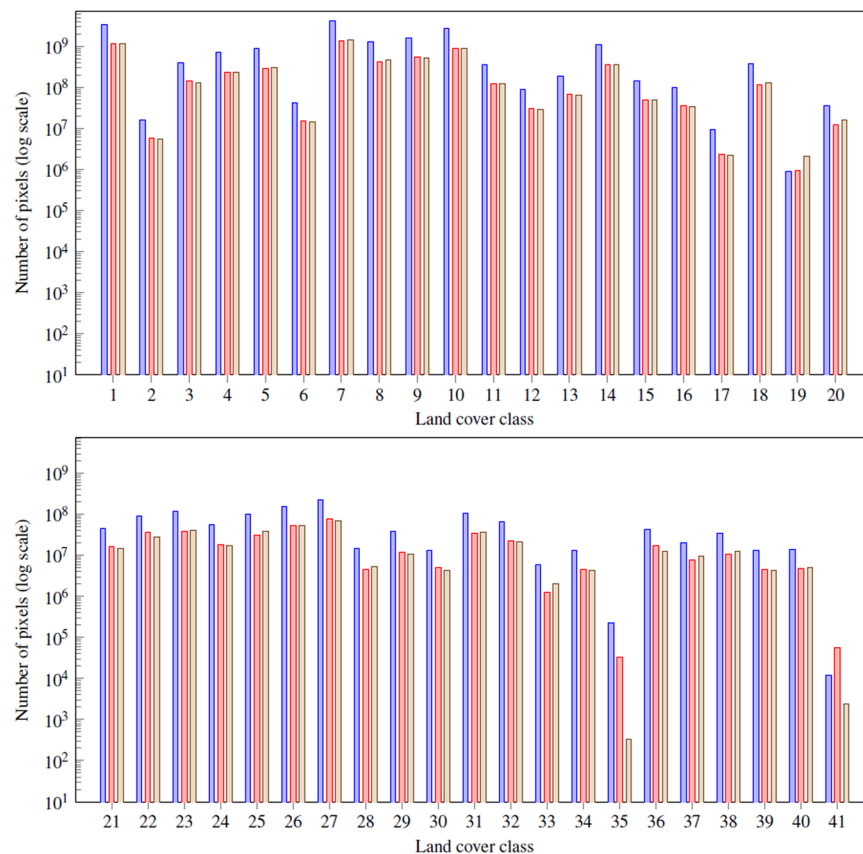


Fig. 14 Distribution of the CatLC dataset in three sets: Blue for the train set, red for the validation set and brown for the test set.

The images were processed with the SNAP (Sentinel Application Platform) software⁵ from ESA using the following procedure:

1. Download of the precise orbit for each image using the “Apply-Orbit-File” function, which provides detailed information for its correct georeferencing.
2. Deletion of noisy pixels from the edge of the image using the “Remove-GRD-Border-Noise” function.
3. Radiometric calibration of each image providing calibrated reflectivity information for the Sentinel-1 images. A correct calibration is necessary for the multitemporal study of the data.
4. Topographic effects Compensation using the “Terrain-Flattening” function. The acquisition geometry of the SAR images is oblique, which generates distorting artifacts in the reflectivity associated with the terrain topography (layover, foreshortening and shadowing). This processing compensates for these artifacts to obtain an image that is as independent as possible from the topography.
5. Georeferencing using the “Terrain-Correction” function and final mosaicking of the images.

A video comparing the average Sentinel-1 image and the image for each month is on the CatLC webpage⁶.

Sentinel-2. Sentinel-2 provides multispectral imagery data at different resolutions approximately every five days. We have selected two relevant dates for this dataset, the first one in April and the second one in August. We have chosen these two dates to follow the phenological evolution of the vegetation throughout the spring and late summer. As we are in the Mediterranean area, with these two dates it is possible to detect both winter and summer herbaceous crops as well as evergreen and deciduous forest areas. Due to the presence of clouds, multiple data takes have been necessary to make a cloud-free mosaic (see Fig. 9).

The images obtained by the MSI sensor from the Sentinel-2A and 2B satellites, from the European Commission Copernicus program, have been atmospherically corrected by means of the ESA sen2cor v2.8 software⁷ to yield Level-2A images.

The main purpose of sen2cor is to correct single-date Sentinel-2 Level-1C Top-Of-Atmosphere (TOA) radiance from the effects of the atmosphere in order to deliver a Level-2A Bottom-Of-Atmosphere (BOA) reflectance. The process may optionally use a DEM (Digital Elevation Model) to correct the changes in the radiometry related to the topographic relief. A 10 m gridded DEM generated at ICGC by photogrammetric techniques has been used in this study. A total of 10 bands, at 10 m and 20 m resolution, are preserved as input features for the Deep Learning process. Figure 10 presents Sentinel-2 images before and after they have been corrected.

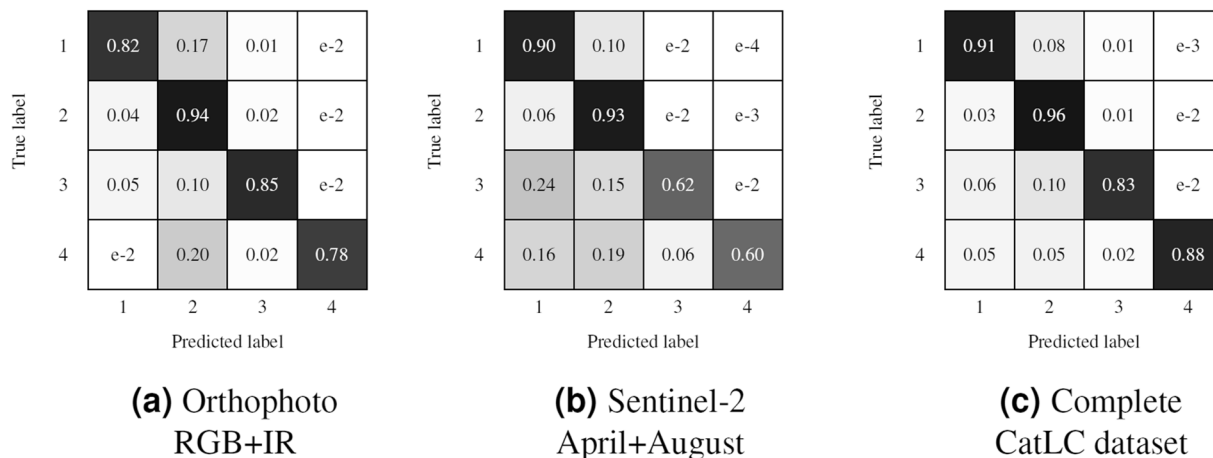


Fig. 15 Confusion matrix using different input data. All trained with U-Net neural network. The 41 classes have been compacted to the 4 superclasses (1: agriculture, 2: forest, 3: urban, 4: water).

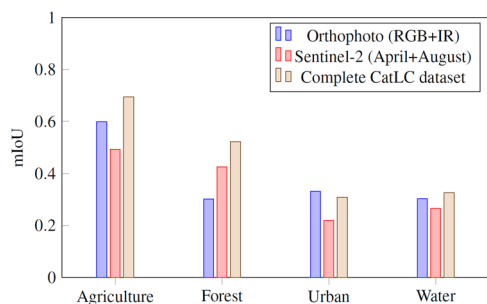


Fig. 16 Mean Intersection over Union using different input data for 4 superclasses.

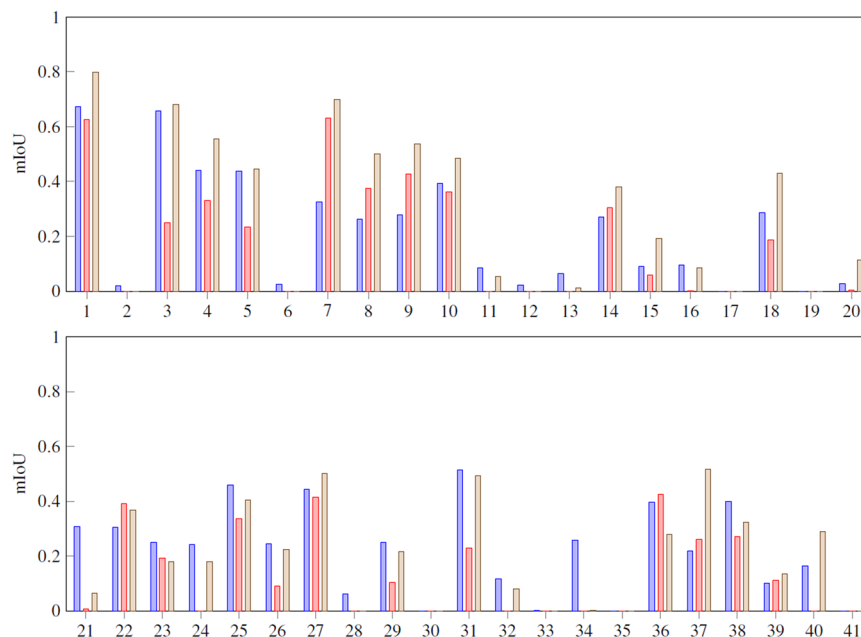


Fig. 17 Mean Intersection over Union using different input data for 41 classes.

Technical Validation

Assessing the quality of the images in the dataset is very important to ensure that the input data is of optimal quality, apart from its characteristics such as the spatial resolution or the number of spectral bands. In order to show the quality of the presented data, below there is a summary of the Quality Controls (QC) used during their processing:

1. Land Cover map: The land cover map update process includes periodic checks of some selected polygons among those that have been geometrically and semantically modified. In the case of systematic errors or misunderstandings about the legend, they are corrected. At the end of the 2018 update, an internal quality supervision has been carried out on a sample of 811 points throughout the territory, resulting in a thematic accuracy of 81%.
2. Orthophoto: According to the specification⁴, several automatic and manual checks are carried out. The main ones are positional accuracy (RMSE 0.5 m), geometric and radiometric continuity, dynamic range and image quality. Additional manual inspection after retouch ensures that remaining artifacts cover less than 1% of the total area of Catalonia.
3. Sentinel-1 and Sentinel-2: The processing of the Sentinel-1 and Sentinel-2 images has been carried out with the quality standards of the ESA SNAP⁵ and sen2cor⁷ software respectively. This guarantees a good radiometric and geometric calibration of the images. After processing, an evaluation of the location of various control points has been performed to validate a geolocation error of less than one pixel (10 m) in absolute value and relative to the rest of the products in the dataset
4. Digital Elevation Model: As stated in⁸ the resulting DEM is checked for positional accuracy (RMSE 0.9 m), logical consistency and completeness.
5. Digital Surface Model and Canopy Height Model: The quality of both the LiDAR Digital Terrain Model (DTM) and the DSM was checked against a high number of independent check points, measured with GPS at a vertical accuracy of approx. 4 cm. In case of the LiDAR DTM, the check points are located on soccer fields. For the DSM the available country-wide network of photogrammetry control points was used. Only points located on the ground were selected (around one thousand points). From the check points the following empirical vertical accuracy values (Root Mean Squared Errors - RMSE) were derived:
 - LiDAR DTM: better than 15 cm.
 - DSM in the Pyrenees: better than 40 cm.
 - DSM in the rest of Catalonia: better than 30 cm.
 - CHM in the Pyrenees: better than 45 cm.
 - CHM in the rest of Catalonia: better than 35 cm.

Since the DSM and the CHM are automatically generated products, their quality can be considerably decreased in areas where the matching algorithm did not achieve optimal results (e.g. in shadow areas).

It should be also noted that in areas covered with some kind of forests and mildly sparse trees the DSM/CHM does not always represent the height of the canopy, depending on the tree density and the presence of foliage.

Usage Notes

An initial benchmark accompanies the CatLC dataset as a starting point and to show a helpful pipeline to train a model with provided data. Note that the use of these data is subject to a Creative Common International Recognition 4.0 license, and contains Sentinel Copernicus data modified by the ICGC.

Unlike other datasets that have multiple images, CatLC has only one large image. To work with it, we will need to access smaller tiles, so the first step has been to create a list with the indexes of all the tiles that we are going to use in the dataset of dimension 960×960 pixels (in the higher spatial resolution images of 1 m). This list was then randomly divided into three groups, 60% for training, 20% for validation, and 20% for testing. Being this a segmentation problem, we have not been able to have a homogeneous distribution of the three groups because usually tiles contain multiple classes. The distribution for the sets can be found in Fig. 14.

As the main baseline, we have selected the classical U-Net neural network¹⁰, which is used as a starting point in most applications that require image segmentation. This baseline has been implemented in PyTorch, running in a workstation with a Nvidia Quadro P5000 GPU. Cross entropy loss has been used, together with an Adam optimizer with 0.0001 as the learning rate.

Experiments consider three different scenarios:

1. Use as input data RGB orthophotos and infrared band. We know by experience that the high resolution should give good results in the border between different classes, but its limitations regarding spectral bands makes it harder to differentiate classes that belong to the agricultural or forest superclasses.
2. Use as input data two Sentinel-2 images corresponding to April and August. This time, the low resolution will penalize the frontiers, but there should be an improvement in differentiating agricultural or forest superclasses.
3. Use as input data the complete CatLC dataset. It does not make sense to use Sentinel-1 or topographical data all alone because most of its information is about elevation or reflectivity. But the combinations of those with orthophotos and Sentinel-2 data should improve the results.

To better visualize the results, we have compressed the data in a four superclasses confusion matrix (Fig. 15) and mean intersection over union metrics (Fig. 16) as recommended in COCO dataset¹¹. In Figs. 17–20 a confusion matrix and a mean intersection over union for all the 41 classes are shown.

As we stated before, Sentinel-2 outperforms the orthophoto in agricultural and forest zones, but it loses when we need more resolution as in urban areas. Finally, using the complete dataset gives better results overall. In Fig. 21 there is an example of a segmentation using the complete CatLC dataset.

Code availability

CatLC is available for download, along with all the necessary information and tutorials, on the following [website](#)⁶.

There is a tutorial on how to manage the data in the following url: <https://github.com/OpenICGC/CatLC/>. There is also the code to reproduce the training presented in the article. We provide the logs for the whole training that can be visualized using Tensorboard.

Received: 22 April 2022; Accepted: 4 September 2022;

Published online: 08 September 2022

References

1. Deng, J. *et al.* Imagenet: A large-scale hierarchical image database. IEEE conference on computer vision and pattern recognition 248–255 (2009).
2. Everingham, M., Gool, L. V., Christopher, K. I., Williams, J. & Zisserman, A. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html> (2012).
3. García, C., Vitrià, J. & Mora, O. Uncertainty-based human-in-the-loop deep learning for land cover segmentation. *Remote Sensing* **12**, <https://doi.org/10.3390/rs12223836> (2020).
4. ICGC Orthophoto technical specifications, <https://datacloud.ide.cat/especificacions/ortofoto-25cm-v4r0-esp-02ca-20190319.pdf> (2019).
5. ESA. Sentinel application platform (snap). <https://step.esa.int/main/toolboxes/snap/> (2020).
6. García, C., Mora, O., Pérez-Aragüés, F. & Vitrià, J. Catalonia multi-resolution landcover dataset (catlc), <https://www.icgc.cat/en/Downloads/Maps-in-image-format/Catalonia-Multi-resolution-Landcover-Dataset-CatLC> (2022).
7. Main-Knorn, M. *et al.* Sen2Cor for Sentinel-2. In Bruzzone, L., Bovolo, F. & Benediktsson, J. A. (eds.) *Image and Signal Processing for Remote Sensing XXIII*, 3, <https://doi.org/10.1117/12.2278218> (SPIE, Warsaw, Poland, 2017).
8. ICGC Dem technical specifications, <https://datacloud.ide.cat/especificacions/model-digital-terreny-5x5-v1r0-esp-02ca-20130601.pdf> (2013).
9. ICGC Catalonia lidar dataset, <https://datacloud.ide.cat/especificacions/model-digital-terreny-2x2-v1r0-esp-01ca-20160519.pdf> (2016).
10. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F. (eds.) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241 (Springer International Publishing, Cham, 2015).
11. Lin, T.-Y. *et al.* Microsoft coco: Common objects in context (2014). Cite arxiv:1405.0312Comment: 1) updated annotation pipeline description and figures; 2) added new section describing datasets splits; 3) updated author list.

Acknowledgements

This research was funded by industrial doctorate grant 2018-DI-0045 of AGAUR between University of Barcelona and Cartographic and Geological Institute of Catalonia and RTI2018-095232-B-C21, 2017SGR1742 grants. We also want to acknowledge the work of photointerpreters and remote sensing experts at the Earth Observation Area (CSPCOT) at Cartographic and Geological Institute of Catalonia. J.V. acknowledges funding from projects No. RTI2018-095232-B-C1 (MINECO/FEDER, UE) and 2017SGR1742 (Generalitat de Catalunya). Finally, we want to thank the contribution of Vicenç Palà, Lydia Pineda, Anna Tardà, Andres Alvarruiz and Wolfgang Kornus to the description of the data quality control.

Author contributions

Carlos García conceived the idea of publishing the dataset and has written a large part of the manuscript. Oscar Mora has contributed to the creation of the dataset and to the writing of the sections referring to Sentinel-1 data, land cover map and topographic maps, as well as the review of the work. Fernando Pérez has contributed to the writing of the sections referring to Sentinel-2 and orthophoto data, as well as to the review of the work. Jordi Vitrià has contributed to the writing and analysis of the Usage Notes, as well as to the revision of the work. All authors discussed the results and contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to O.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022