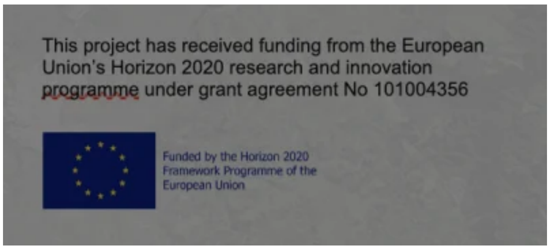
Land Use Land Classification (LULC) Change Detection with High Cadence Multimodal Image Time Series and Self-Supervised Learning.

Akhil Singh Rana-1, Caglar Senaras-1 , Benjamin Bischke-3, Patrick Helber-3, Timothy Davis-1, Wanda Dekeersmaecker-4, Daniele Zanaga-4, Annett Wania-1, Ruben Van De Kerchoveand-4, Giovanni Marchisio-2

1 Planet Labs GmbH, Germany. 2 Planet Labs PBC. 3 Vision Impulse GmbH and DFKI, Germany. 4 VITO NV, Belgium.



PRESENTED AT:

INTRODUCTION

We believe that the future of Earth Observation (EO) is in fusion, harmonization, and interoperability of satellite imagery. Intensified monitoring leads to better understanding of land use and reduction of maintenance costs for all Land Cover products.

RapidAI4EO is an initiative that aims to establish the foundations for the next generation of Copernicus Land Monitoring Service (CLMS) products. The goal is to provide intensified monitoring of Land Use (LU), Land Cover (LC) changes at a much higher spatial resolution and temporal cadence than is possible today. Key objectives are to explore, evaluate, and quantify state of the art deep learning algorithms and methodologies that leverage three meter, daily time series, in conjunction with higher spectral resolution Sentinel-2 imagery.

Consortium Partners:
The RapidAI4EO projects brings together Planet Labs PBC, the operator of the world's largest fleet of Earth-imaging satellites and the recognized leader of the CubeSat revolution, VITO, the main production center of the Copernicus Global Land Service, Vision Impulse, a recent spin-off of German Research Center for Artificial Intelligence (DFKI, the largest research center for Artificial Intelligence in the world and one of the two European NVIDIA AI Labs), the International Institute for Applied Systems Analysis (IIASA) whose Center for Earth Observation and Citizen Science (EOCS) devises new approaches and technologies to collect data on land cover and land use, and Serco Italia, a worldwide service provider to governments, international agencies and industries, and operator of the ONDA DIAS platform.

The objectives of RapidAI4EO are:

1. the creation and release of the most comprehensive spatiotemporal EO training sets ever produced for machine learning;
2. the development and implementation of novel AI solutions for continuous change detection that leverage these data sets;
3. the ability to drive frequent temporal updates of the Corine Land Cover (CLC) product; and
4. to demonstrate improved LULC mapping using harmonized Sentinel-2 and very high resolution, high cadence data streams.

MACHINE LEARNING

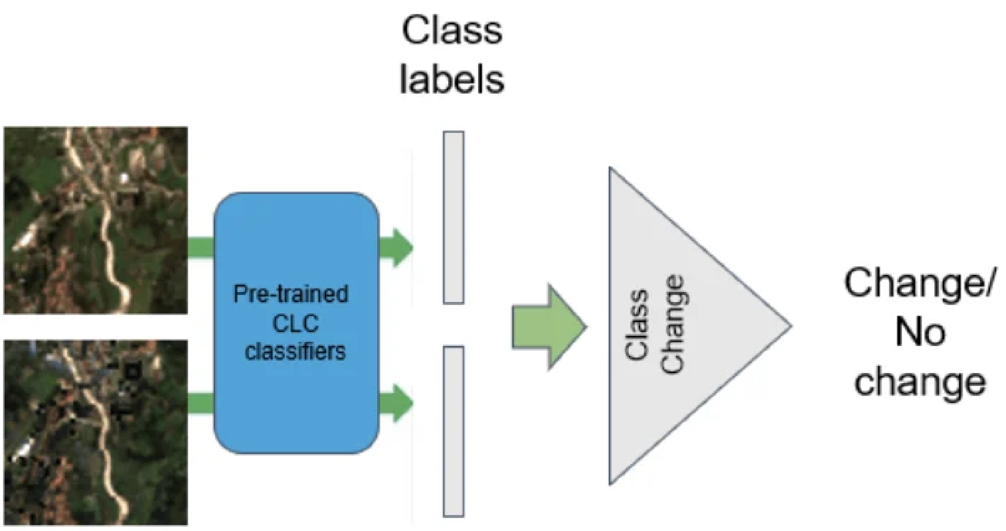
The time-series datasets provide a wealth of information from which machine learning models can be trained for both classification and change detection. The timeseries image cubes will drive innovation in change detection modeling, which is a core research area of the RapidAI4EO project.

Multiple change detection models are being constructed and compared on the dense Planet Fusion data as well as the sparse Sentinel-2 time series. A comparison will provide insights as to what spatial, radiometric, and temporal resolutions are required to satisfactorily detect changes. Ultimately, these various models will be integrated to produce a heat map of change, which is the basis for identifying areas where the CLC is likely to require updating and ultimately producing the inventory on a shorter cycle than has historically been possible.

We work on, and leverage from, two styles of machine learning model training: supervised learning and unsupervised learning.

Supervised learning:

Supervised learning is a technique where models learn their parameters from a labelled training dataset.
Supervised learning in general produces better results than unsupervised learning if a labelled dataset is available, clean and covers the problem set quite well. This is not always the case.
In our case, CLC labels for 2018 are available to train our supervised models. Feature extractors can be trained as a simple multi-label classifier, and then perform change detection based on changes in class labels as illustrated in the figure below.



Change Detection approach with class labels

Some of the initial findings of supervised learning already show improved accuracy with planet fusion(PF) over Sentinel-2 as shown below. This result can be explained by better spatial resolution of PF over Sentinel-2 or because of improved imagery i.e. clouds free and less artifacts, because of our fusion technique.

Comparison Sentinel-2 vs. PlanetScope Fusion

Method	Bands	Satellite	Months	Classes	F1-Score
ResNet-50	All	S2	All months	44 (L3)	65.66
ResNet-50	All	PS	All months	44 (L3)	**68.92**

Method	Bands	Satellite	Months	Classes	F1-Score
ResNet-50	All	S2	All months	15 (L2)	74.24
ResNet-50	All	PS	All months	15 (L2)	**77.30**

Conclusion:

- Better CLC classification accuracies with PlanetScope Fusion compared to Sentinel-2

Multi-Label Classification: RapidAI4EO 44 classes and 15 classes, F1-Score (micro), 20 epochs, Multispectral images as input (no timeseries!)

RapidAI4EO

Initial findings also point to the fact that temporal models can outperform mono-temporal models, i.e. LSTM vs CNN, as shown below.

Method	Bands	Satellite	Months	Classes	F1-Score
ResNet-50	All	S2	All months	15	74.24
ResNet-50	All	PS	All months	15	77.30
LSTM	All	PS	All months	15	80.10

Unsupervised learning:

Unsupervised learning in machine learning is the domain where ML engineers need not prepare a labelled training dataset or supervise the model training. This approach instead lets the model figure out the optimal solution to the problem on its own.

Learning without supervision can be more unpredictable compared to supervised learning. However, it would be prohibitively laborious to prepare a labelled training dataset for change detection, for 500,000 spatial locations and daily temporal data of two years. Therefore, it becomes even more important, challenging and exciting to benefit from unsupervised learning for this task.

We are investigating the following unsupervised deep learning approaches:

1) autoencoders

2) disentangled training with Siamese-like networks

3) student-teacher architectures, like BYOL(Bootstrap Your Own Latent), where we prepare positive and negative training pairs for our model, and

4) Vision Transformers such as DINO.

Subsequent results presented in this poster focuses on DINO.

Many of the most recent exciting breakthroughs in deep learning have come from two recent innovations: self-supervised learning, which allows machines to learn from random, unlabeled examples; and transformers, which enable AI models to selectively focus on certain parts of their input and thus reason more effectively. DINO (https://arxiv.org/abs/2104.14294) is a Vision Transformer (ViT) which can be trained with no supervision.

We have slightly modified the dataset and model architecture to make it suitable for remote sensing data.

[VIDEO] https://res.cloudinary.com/amuze-interactive/image/upload/f_auto,q_auto/v1638725547/agu-fm2021/c5-06-3b-d7-07-b5-95-ca-81-ca-dd-db-fa-66-15-49/image/dino_dcgqdt.mp4

Figure source: github (http://Figure source: https://github.com/facebookresearch/dino)

Unlike some of the self-supervised methods which require both negative and positive pairs to learn, Dino needs only positive images. Self-supervision of DINO is done by using student-teacher architecture. The parameters of the teacher network are learnt with an exponential moving average method, and it learns from the student network. Dino architecture mainly consists of transformers or residual blocks as the main backbone for feature representation learning. The output of these feature backbone goes into attention heads. A Transformer module in its simplest definition, is a combination of attention blocks and fully connected feed-forward layers. More about the architecture can be learnt in the official DINO paper (https://arxiv.org/abs/2104.14294), as it is out of the scope for the work presented.

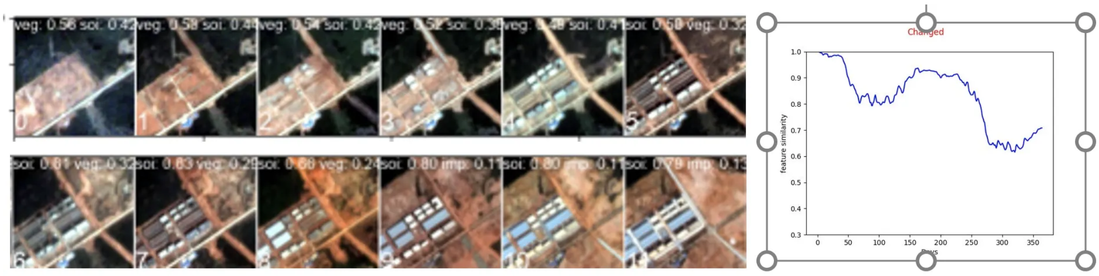
To make DINO work with our dataset and remote sensing in general, we did following adaptations to the original codebase:

- We adapt the input layer of DINO to accept 4-channel, RGB-NIR, bands of Planet fusion data.
- Individual bands of our data are normalized with pre-computed mean and standard deviation for each channel(RGB-NIR) in our data.
- In order to select positive training samples, we select local and global crops as a list of images from the same area of interest AOI randomly selected within the range of [15-30 days], for local crops. Whereas, global crops are images of the same AOI in the time range [30-60 days]. This helps self supervision and student-teacher networks to learn a robust representation against seasonal and phenology changes, and learn a representation of CLC classes within a 200x200 patch timeseries.
- The architecture was modified to get features from intermediate attention blocks and final learnt features are retrieved as a vector. This feature vector/embedding is then used for the downstream task of change detection. Features are passed to the change detection module, where structural changes are detected. We expect that the features that DINO is learning, learn to ignore phenology and noise in our data.

CHANGE DETECTION

Once our feature embedding models are trained, the second part of our work is to use them in order to detect changes. As already mentioned, for supervised models we can simply compare changes in CLC classes between different timesteps.

For our unsupervised model (DINO), we need to implement additional refined analyses like performing clustering in feature embedding space to separate similar embeddings from those that reflect real structural change. Another, simple way is to compare feature embeddings and compute cosine distances, then perform thresholding to detect a change. This approach is used for the results that are shown in the figure below.

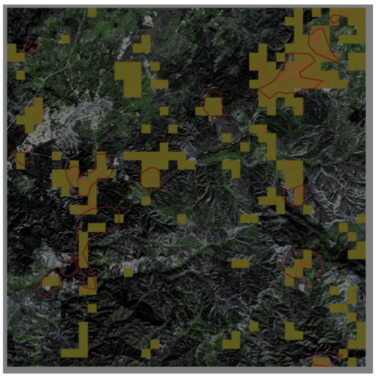


Selecting a threshold value for cosine similarity (1- cosine distance) can impact the results. Different threshold values, mean different sensitivity to detect amount of change in the patch, a high value means a very little change in pixels can trigger the module to detect changes, whereas a higher threshold value will only detect bigger patch change. This also means a low value has higher chance of detecting more false positives. Therefore, it is important to analyze a few test areas before making a decision on the final threshold value.

This simple distance based approach, in cosine space of learnt high dimensional features can be used to produce heatmaps of big regions as described in the results section.

RESULTS

We took one of the validation AOIs from Portugal to generate a change heatmap. As shown below. AOI is divided into grid size of 600 x 600 meters, the patches go to our trained machine learning feature extractor(DINO, for this result), and cosine distance is computed in feature space to mark changes in 200x200 patches.



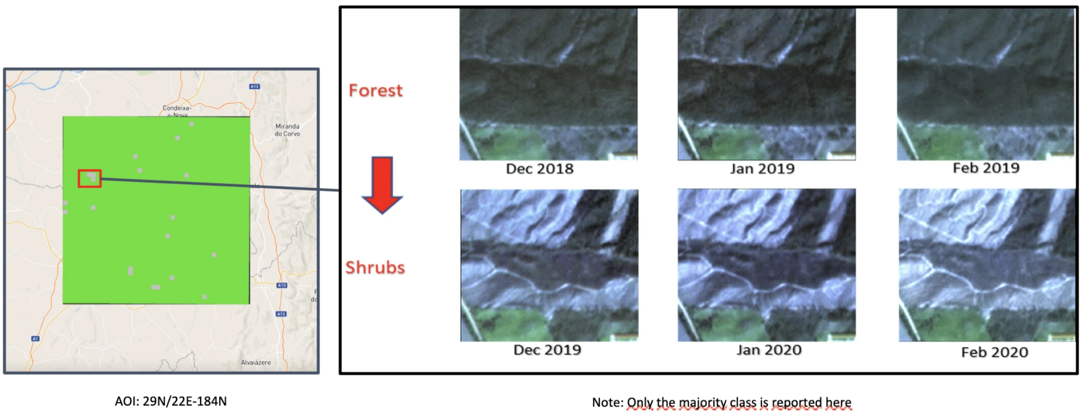
Steps to get the final heatmap:

- Use feature embedding models trained either in supervised or unsupervised fashion (in this case we show results from DINO).
- Generate feature embeddings for two years.
- Compute the feature similarity or cosine distance.
- Apply thresholding to generate daily heatmaps and compute mean yearly heatmap from it.

The animation below shows the daily heatmaps. There are many false positives and outliers in the daily change heatmaps. This can be explained by different phenelogy in the imagery or can be a false alarm from our models. Therefore, it is important to ensemble these heatmaps into a mean heatmap, which helps to ignore all the subtle changes which are not consistent over the time frame of a year. This is another benefit of using higher temporal information.

[VIDEO] https://res.cloudinary.com/amuze-interactive/image/upload/f_auto,q_auto/v1639327164/agu-fm2021/c5-06-3b-d7-07-b5-95-ca-81-ca-dd-db-fa-66-15-49/image/heatmap_rescaled_256_point_vq9z41.mp4

We show below an example where a CLC class changes from forest to shrub and was identified by our model.

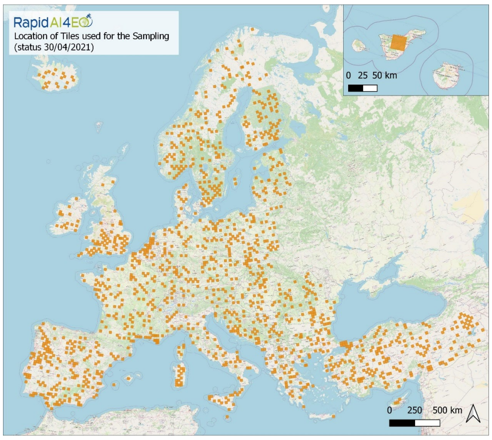


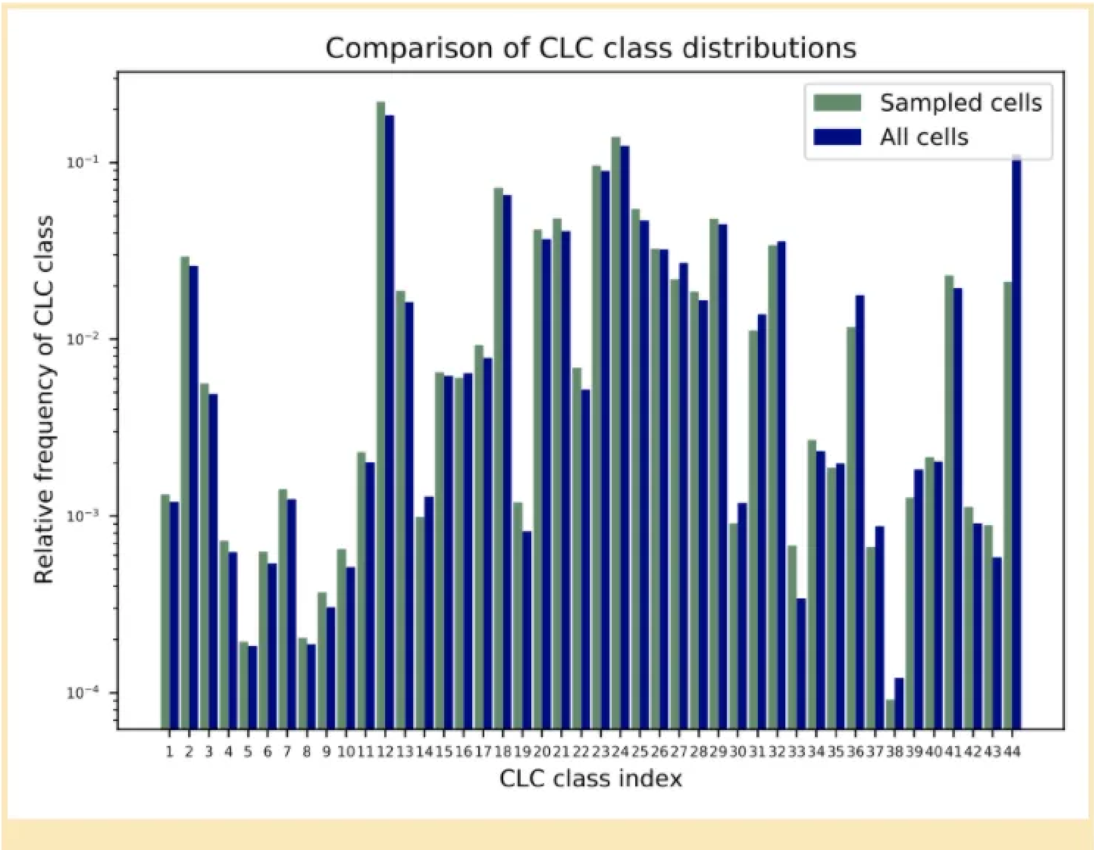
[VIDEO] https://res.cloudinary.com/amuze-interactive/video/upload/vc_auto/v1638626777/agu-fm2021/C5-06-3B-D7-07-B5-95-CA-81-CA-DD-DB-FA-66-15-49/Video/Screen_Recording_2021-12-04_at_19.00.51_z3acjq.mp4

DATASET

Overview

The inspiration for the RapidAIO4EO corpus comes from the BigEarthNet (https://ieeexplore.ieee.org/abstract/document/8900532) and Eurosat (https://ieeexplore.ieee.org/abstract/document/8736785) datasets and is based on the notion of image patches. The main difference is that we provide high-cadence time series of images at all locations to better capture land cover dynamics. The 500,000 sampling locations are carefully selected locations throughout the European continent to account for country representation and CLC class distribution as shown below.





Description of CLC index to class mapping can be found on this document (https://land.copernicus.eu/eagle/files/eagle-related-projects/pt_clc-conversion-to-fao-lccs3_dec2010). And overview can be seen in the table below classes

idx	label		
1	Continuous urban fabric	15	Vineyards
2	Discontinuous urban fabric	16	Fruit trees and berry plantations
3	Industrial or commercial units	17	Olive groves
4	Road and rail networks and associated land	18	Pastures
5	Port areas	19	Annual crops associated with permanent crops
6	Airports	20	Complex cultivation patterns
7	Mineral extraction sites	21	Land principally occupied by agriculture with significant areas of natural vegetation
8	Dump sites	22	Agro-forestry areas
9	Construction sites	23	Broad-leaved forest
10	Green urban areas	24	Coniferous forest
11	Sport and leisure facilities	25	Mixed forest
12	Non-irrigated arable land	26	Natural grasslands
13	Permanently irrigated land	27	Moors and heathland
14	Rice fields	28	Sclerophyllous vegetation
29	Transitional woodland-shrub	36	Peat bogs
30	Beaches dunes sands	37	Salt marshes
31	Bare rocks	38	Salines
32	Sparsely vegetated areas	39	Intertidal flats
33	Burnt areas	40	Water courses
34	Glaciers and perpetual snow	41	Water bodies
35	Inland marshes	42	Coastal lagoons
43	Estuaries	45	NODATA
44	Sea and ocean		

Data Sources

Imagery in the RapidAI4EO dataset is drawn from two sources: Sentinel-2 and the PlanetScope constellation. The image patch size for each of the 500,000 locations in the RapidAI4EO training corpus is 600m x 600m. The time series covers two full years: 2018 and 2019. We provide multi-class label annotations showing the number, types and percentage of land cover types at all patch locations based on the CLC 2018 product.

Sentinel-2 is a multispectral earth observation mission of the European Space Agency (ESA). Consisting of two satellites, this mission acquires imagery with 12 spectral bands in the visible and infrared range. The ground resolution is 10–60 meters, depending on the spectral band. Although Sentinel-2 has a revisit time between three and five days on average, many observations may contain clouds. To deal with such effects, we produce time series with minimal cloud interference by fusing images from different time steps per month. We use atmospherically corrected Sentinel-2 Level 2A products and process all 12 available multispectral bands. As a result, each of the 500,000 sites is represented by a time series of 12 Sentinel-2 images.

PlanetScope is a constellation of 120 cube satellites owned and operated by Planet Labs PBC. These satellites capture imagery at nadir in four spectral bands from the visible and near-infrared (VNIR) range, with a ground resolution of up to three meters. The second set of time series is composed of Planet Fusion images which have been harmonized with Sentinel-2. The Fusion process produces a daily time series of cloud-free images by applying the cubesat enabled spatio-temporal enhancement method (CESTEM, Houborg and McCabe, 2018 (https://www.sciencedirect.com/science/article/abs/pii/S0034425718300786)). CESTEM begins by harmonizing PlanetScope imagery against that of MODIS, Sentinel-2, and Landsat 8 to produce output with PlanetScope's three meter ground resolution and daily cadence, and the radiometric consistency of the reference sensors. CESTEM also includes a sophisticated approach to gap-fill pixels that are obfuscated by clouds, cloud shadows or other atmospheric phenomena, or in case no acquisition data are available. Gap-filling is done with clear-sky observations on proximate days, as well as from the same time period in previous years. CESTEM applies a combined machine learning and regression approach trained on available observations to predict pixel values that require gap-filling. Quality assurance layers are also generated as part of the process to provide insight as to whether pixels were gap filled and if so how much information was available to inform the process. These layers allow users to make informed decisions about the confidence of the gap-filling result and whether they would like to implement any masks.

The multi-temporal image sequence at each location enables the development of novel LULC image classification and change detection models.

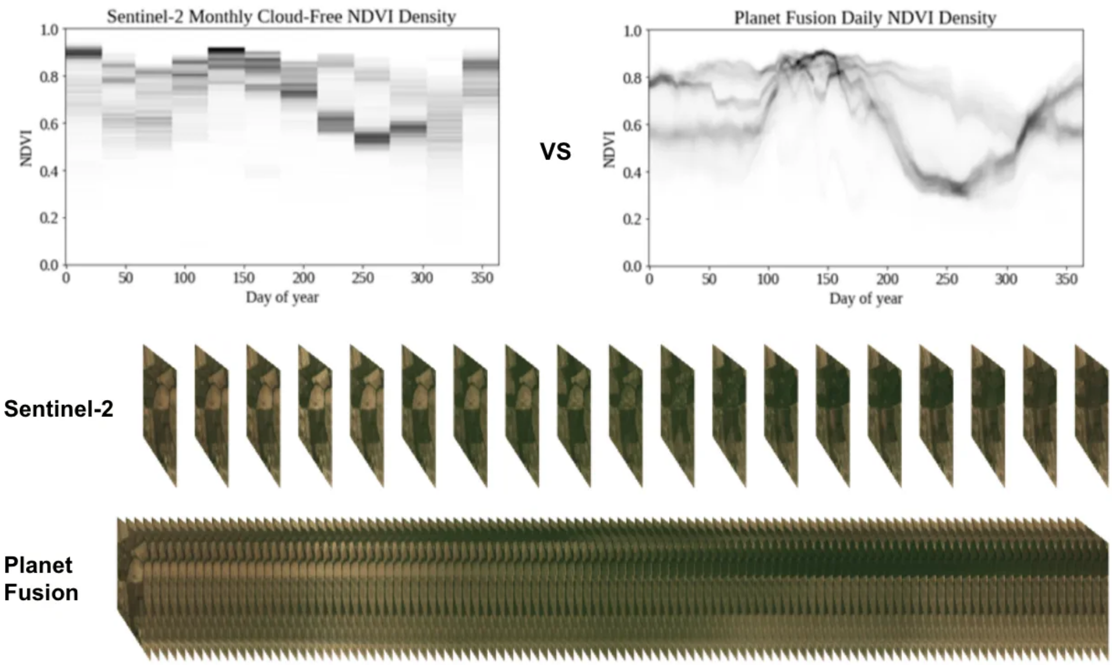
Overview of the two satellite constellations used to derive the time series image cubes for the dataset.

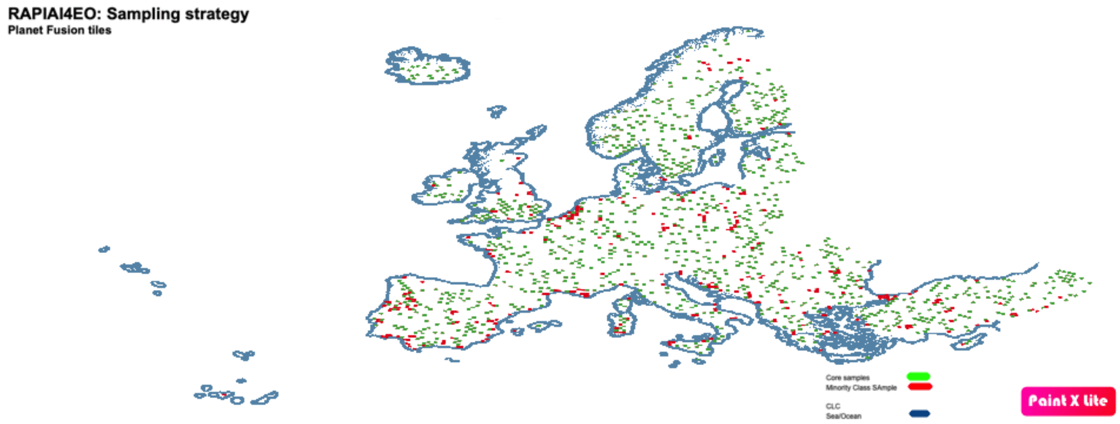
	Sentinel-2	PlanetScope
No. satellites	2	120
No. bands	12	4
Revisit time (days)	3–5	1
Ground resolution (m)	10–60	3–4
Patch size (pixels)	60×60	200×200

Dataset Highlights:

- 600m x 600m patches with Sentinel-2 and Planet Fusion yearly time series
- Sampled at 500,000 locations, for two years, accounting for:
  - CLC class distribution
  - Country representation
- Multi-class annotations based on CLC 2018
- Designed for LULC classification and change detection use case, but generalisable to other problems
- Open sourcing in July 2022

[VIDEO] https://res.cloudinary.com/amuze-interactive/video/upload/vc_auto/v1637946635/agu-fm2021/C5-06-3B-D7-07-B5-95-CA-81-CA-DD-DB-FA-66-15-49/Video/33N_19E-169N_06_08_ndvi_video_rgb_iyagh2.mp4




RAPIAI4EO: Sampling strategy
Planet Fusion tiles
Core samples
Minority Class SAmple
CLC
Sea/Ocean
Paint X lite

FUTURE WORK

The project will continue for 15 months, with the dataset being open sourced for researchers by July next year.

In the immediate future, we plan to combine change heatmaps coming from both supervised and unsupervised feature extractors using model ensemble techniques to further improve our results and reduce false positives.

You can follow our progress on the RapidAI4EO website (https://rapidai4eo.eu/).

AUTHOR INFORMATION

Akhil. S. Rana[1], Caglar Senaras[1], Benjamin Bischke[3], Patrick Helber[3], Timothy Davis[1], Wanda Dekeersmaecker[4], Giovanni Marchisio[2], Daniele Zanaga[4], Annett Wania[1], Ruben Van De Kerchoveand[4]

[1] Planet Labs GmbH, Germany.

[2] Planet Labs PBC.

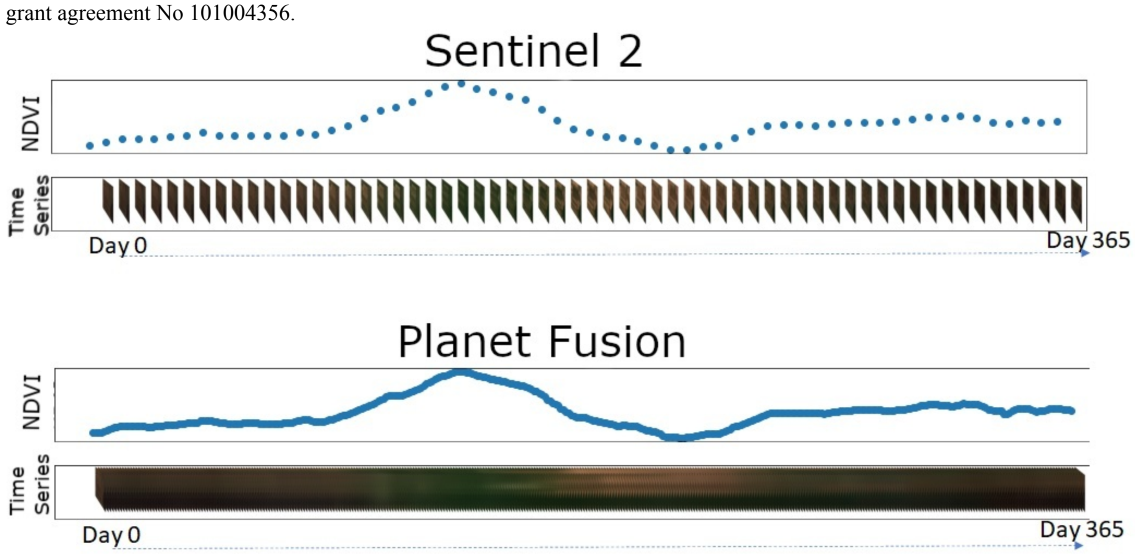
[3] Vision Impulse GmbH and DFKI, Germany.

[4] VITO NV, Belgium.

ABSTRACT

RapidAI4EO is an initiative that aims to establish the foundations for the next generation of Copernicus Land Monitoring Service (CLMS) products. The goal is to provide intensified monitoring of Land Use (LU), Land Cover (LC) changes at a much higher spatial resolution and temporal cadence than it is currently possible. Key objectives are to explore, evaluate, and quantify state-of-the-art deep learning algorithms and methodologies that leverages 3-meter daily time series, in conjunction with higher spectral resolution Sentinel-2 imagery. This data set which includes 500,000 locations across Europe from the year 2018, will be open-sourced for the benefit of the remote sensing and machine learning community. The cloud-free, harmonized, high spatial and temporal resolution time-series satellite data is created using a data fusion methodology (CESTEM: CubeSat Enabled Spatio-Temporal Enhancement Method) that Planet has been pioneering. While the daily cadence presents a challenge from the point of view of manually labeling GT for changes, it also offers an opportunity for the development of new unsupervised models that can learn how to disentangle the phenology signal from proper structural changes. In order to achieve this goal, we are developing different machine learning techniques, including supervised, self-supervised, and unsupervised approaches. For our supervised approach, we use existing CORINE Land Cover (CLC) labels to train a patch classifier which will be used to detect the class changes in the time series. We also explore self-supervised and unsupervised deep learning methods, such as contrastive learning, variational autoencoders, etc. We then use the learned features from these approaches for a downstream task of change detection and compare their performances.

*RapidAI4EO has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101004356.



(https://agu.confex.com/data/abstract/agu/fm21/7/1/Paper_886717_abstract_820404_0.jpg)