Rules of the competition

1. Organiser

Leaf Nothing Behind is a competition organised non-professionally by the physical person Jules Salzinger under the informal name **Transfer Learning**. The Competition is available at https://transfer-learning.org/competition and it is organised remotely.

2. Duration and Deadlines

The deadline to register is **2023, April 7th 11:59PM (Paris Time)** and the competition itself will take place from **March 30th 8:00PM (Paris Time) to April 20th 11:59PM (Paris Time)**, which is the final deadline for submissions.

3. Conditions of access to the game

Can participate in this competition any natural persons (Master's students, PhD students, students with knowledge of data science, and data professionals) with working proficiency in English.

Participants have to register through a team of a **maximum of 4 participants**. If all participants in a team rightfully claim to be first year students in one of the three schools CentraleSupélec, École Centrale de Lille or École Centrale de Marseille, they will be eligible to compete for the so-called "beginner leaderboard" among other such groups. Until April 7th, they may request to be moved to the "advanced leaderboard" instead. All teams that are not eligible to compete for the "beginner leaderboard" will by default be registered in the "advanced leaderboard".

Participation is free of charge.

4. Terms of the Competition

The registration portal is located at https://competition.transfer-learning.org/register .
Groups are **required** to provide the following information:
Number of participants and for each participant:
- **Full name**
- **Valid email address**

- **University of affiliation** and **degree or year of study**, when applicable
- **Company** and **job position**, when applicable

The participants need access to a computer device to take part in the competition. Participation in the competition implies unreserved acceptance of current regulations.

Goal of the competition

Context

Multi-spectral satellite imagery is a very valuable source of information about earth and human activities. It can be used for farming, environmental monitoring, planning, and many other tasks at the core of our society's inner workings. However, multi-spectral imagery suffers from cloud occlusions. This makes it trickier to leverage information from this data, as several different timestamps and perspectives need to be somehow aggregated to gather a reliable view of a given area at a given time. In this competition, we decide to tackle the following research topic.

The considered satellite imagery is the [Leaf Area Index](#) obtained from [Sentinel-2 imagery](#). Sentinel-2 covers any monitored area every few days (2 to 3 days in Europe depending on latitude) and gathers precise, but lossy multi-spectral data. In order to get additional information on the location at a given timestamp, we also consider [Sentinel-1 imagery](#). Sentinel-1 collects radar data which is robust to weather conditions, making it a valuable source of conditioning information. Given this data, we focus on the problem of Leaf Area Index prediction at a given timestamp conditioned on previous Sentinel-2 and Sentinel-1 observations.

General problem definition

This competition tackles multi-regression of pixel colours in a 256*256*1 image. Given some context information, participants are tasked with predicting Leaf Area Index values of the next image in a time series. Therefore, a given test sample consists in :
- some general metadata
- some 256*256*n pixel-aligned conditioning information C

For training purposes, participants will be given a dataset of time series with information about the location of potential clouds. These time series can be used to the participants' discretion such that the resulting model is capable of inference on the test data.

Participants will submit a model and inference code, which will be subject to constraints in terms of execution hardware and time.

Inferred images will be evaluated based on a custom metric computed between the ground truth and the inferred Leaf Area Index. This metric will make sure that the trained model does not trivially reproduce the last timestep in the time series. Indeed, between two images of a series, few changes are typically observed which can lead a model to converge to trivial solutions. For this reason, our metric will favour models that are able to predict a change when a change actually occurs, which can be considered a rare occurrence.

Data description

A detailed description of the dataset is available at
https://storage.googleapis.com/laica-training-v3/readme.html

Dataset information

The provided training dataset consists in 22.905 Sentinel-2 samples and 22.905 Sentinel-1 samples (~57 GiB total) collected across 13 unique locations in Europe between 2018 and 2020. In particular, it contains samples offering a good coverage of the different periods in a calendar year. Huge images corresponding to those locations are divided into grids of small 256*256 images which constitute our data. Images within a grid are thus, of course, strongly correlated. The test set contains similarly collected data and consists in samples from locations both present in and absent from the training set. Models are expected to be able to generalise to new images within grids present at training, but also to new uncorrelated areas altogether.

General sample metadata

Metadata on each sample include :
- the location of the sampled data (where the grid is located)
- the location of the image in the grid
- the dates of the collection for each timestep (typically t-2, t-1 and t) and the season considered

Conditioning information

For a given image to predict I at timestep t, the following pixel-aligned conditioning information is given :
- for each of the timesteps t-2, t-1 and t, the radar data from Sentinel-1 is given. At each timestep, this consists in one (256*256*2) image. The two channels represent two bands of the radar data : the cross-polarisation (denoted VH) and the co-polarisation (denoted VV) ;
- for each of the timesteps t-2 and t-1, the Leaf Area Index extracted from Sentinel-2 data is given. At each timestep, this consists in one (256*256*1) image. The one channel represents the extracted potentially incorrect Leaf Area Index ;
- for each available Sentinel-2 image (including ground-truth images), segmentation masks are given. These masks provide additional information that can be used to determine where the provided Leaf Area Index is likely to be incorrect. The segmentation classes are 0 (NO_DATA), 1 (SATURATED_OR_DEFECTIVE), 2 (CAST_SHADOWS), 3 (CLOUD_SHADOWS), 4 (VEGETATION), 5 (NOT_VEGETATED), 6 (WATER), 7 (UNCLASSIFIED), 8 (CLOUD_MEDIUM_PROBABILITY), 9 (CLOUD_HIGH_PROBABILITY), 10 (THIN_CIRRUS) and 11 (SNOW or ICE). Those masks are noted to not always be 100% accurate, but should be sufficient for the task.

All this conditioning information is organised in 256*256 channels pixel-aligned with image I, which is to say that a given pixel location in all those images can be considered information about the same physical location.

The interesting classes to consider in the segmentation masks are :
- 0, 1 and 7, which denote missing or faulty data
- 8 and 9, which denote cloudy area where the Leaf Area Index provided is likely to be incorrect

It should be noted that due to an effect known as the speckle, Sentinel-1 data is typically grainy.

Submission procedure

The exact process to build a correct submission is available in the submission guide : https://drive.google.com/drive/folders/1Z5OydVq2GWkFSbqAUbm5mYAem1wU2DJL?usp=share_link.

Submissions have to be made in the team space of the competition platform : https://competition.transfer-learning.org/my-team.

Evaluation criteria

The quality of the predictions obtained using the submitted code will be assessed using an undisclosed method which will take into account the following considerations :
- submitted methods should accurately predict Sentinel-2 data at time t while only having access to Sentinel-2 data at time t-1 and t-2, corresponding masks and Sentinel-1 data at time t, t-1 and t-2
- submitted methods should be able to generalise to different areas
- submitted methods should be able to generalise to different seasons and weathers
- submitted methods should be able to generalise to situations where significant changes occurred between the Sentinel-2 data at time t-1 and t

Leaderboards

There are two leaderboards in this competition:

- A "beginner Leaderboard", in which only first year students in partner schools can compete
- An "advanced Leaderboard", in which anyone can compete

Information provided by participants about their school year is being checked by the organisers.

Submissions

Submissions have to be done by uploading a zip file on the team space of the competition platform, accessible via the website https://competition.transfer-learning.org/login. Teams should refer to the submission guide (see below) to verify whether their submission

complies with the requested format. Any submission which does not follow the desired format will be ignored during the evaluation. Organisers reserve the right to contact teams when their submission does not run as intended, and to allow them to provide additional information, but such a decision will only be made in exceptional cases and is in no way a prerogative of the participating teams. At the submission deadline, the very last submission uploaded to the team space will be the one considered. Submissions started after the deadline will not be considered in the evaluation.

A submission guide is available at https://drive.google.com/drive/folders/1Z5OydVq2GWkFSbqAUbm5mYAem1wU2DJL?usp=share_link to guide the teams through the submission process.

Evaluation of submissions

The evaluation of the submissions happens in two stages: a mid-competition evaluation and a final evaluation.

First stage - mid-competition evaluation

At a certain, yet to be determined date during the competition, all existing submissions will be evaluated and ranked according to the method which will later be used for the final evaluation. The criteria, data and metrics used will be similar as the ones used for the final evaluation. Results will be communicated to all teams when they are available.

This will serve as a test run for the subsequent final evaluation and an indicator of current performance for the teams.

Second stage - final evaluation

During the time between the submission deadline on April 20th and the rewards ceremony on May 5th, the final evaluation will take place. For each team having submitted a result, the last submission only will be considered. It will be used by the organisers as explained in the submission guide to perform predictions on a private test set, and those predictions will then be analysed using a metric. This metric will serve to rank all teams according to their performance, and the first teams in each leaderboard will then be awarded a prize. Representatives of the first teams in each leaderboard will be permitted to present their method during the rewards ceremony.

Winners announcement

The winners will be announced during the Transfer Learning session on May 5th in the virtual award ceremony. Teams selected for the second stage of the evaluation are recommended to attend this event to present their method if they are announced to be the winners of the competition.

Prizes

There is a total of 300€ awarded to said winning submissions. These are split into:

- the best team in the "beginner leaderboard" will be awarded 100€
- the best team in the "advanced leaderboard" will be awarded 200€

The prize is not transferable. In particular, in the event that a leaderboard would contain no team in their rights to claim a prize according to these rules, the unclaimed prize would remain unallocated.

Both winning teams will have the possibility to **present their approach** at the 14th edition of the Transfer Learning event (the one during which winners are announced), and **their performance will be highlighted** using Transfer Learning's means of communication.

Publication of results and Cession of rights

The aim of the current competition being to advance the knowledge of the scientific community with respect to LAI prediction, winning submissions are meant to be made public under the most permissive license possible given the tools used by the winners.

The proper formalization of the cession of rights allowing such publication is available at https://drive.google.com/file/d/156CQYvpz6ARe1ic6gLT-eG7m5tqvyXAD/view?usp=share_link. The acceptance of the current rules implies the acceptance of this other document.

The participants also agree to publish with the competition organisers a scientific paper describing their approach, as well as the metrics used during the competition. Such a scientific paper will be co-authored by all team members (equal contributions as first authors) and competition organiser Jules Salzinger, who will bring the context as well as the evaluation procedure to the publication. Team members agree to provide all the necessary information to competition organisers, as well as the codebase necessary to reproduce their results, so that they can write formally the content of the paper and provide reproducibility of the work. If they want, winners can also participate in the writing of the

publication. In case no submission reaches a sufficient relevance level, the organisers have the right to give up on this publication.

5.Modification of the competition

The organiser reserves the right, at any time and without notice or obligation to justify its decision, to shorten, extend, postpone or cancel the competition as well as to modify all or part of the conditions of access and/or the terms of implementation of the game. Each potential change will lead to the emission of an email to each contestant explaining it.

6.Personal data

The organisers process your personal data in accordance with the General Data Protection Regulation. You can read more about how your data is processed at https://transfer-learning.org/data .

Personal data will not be kept later than August 1st 2023 unless you have provided us with separate consent. In particular, members of winning teams agree for this data to be kept for up to a year after the end of the competition for publication reasons. Damage resulting from the use of personal data of the participant by unauthorised persons can under no circumstances be passed on to the organisers when they have taken appropriate security measures against such risks.

7.Disputes

All disputes concerning the application of these rules are managed independently by the organisers. In case of a dispute, please send an email with the complaint to transferlearning.event@gmail.com. The reasons for a complaint must be clearly stated, and justified. Complaints will be studied and appropriate actions will be taken within 24 hours, and related information will be addressed by mail at the addresses provided by the participants themselves.

8.Contact

For questions about the competition, please contact: **transferlearning.event@gmail.com**.