Using Deep Learning Methods for Plastic Litter Detection from Satellite Remote Sensor

ESR- Executive Summary Report

Issue 1.0

Date 13 July 2021

Ref.: DL4PlasticLitter_ESR_CG

ESA contract no. 4000131234/20/NL/GLC

Project name: **Using Deep Learning Methods for Plastic Litter Detection from Satellite Remote Sensor**

Document title: **ESR – Executive Summary Report**

Reference no.: **DL4PlasticLitter-ESR-CG**

Issue date: **13-07-2021**

Issue and revision: **1.0**

ESA contract no.: **4000131234/20/NL/GLC**

Organization: **Capgemini, 109 av Eisenhower, 31000 Toulouse**

	Name	Company or Institute	Signature
Prepared by	Delphine Nobileau		
Reviewed by	Delphine Nobileau		
Distribution list	Leopold Summerer Paolo Corradi	ESA/ESTEC ESA/ESTEC	

EUROPEAN SPACE AGENCY

CONTRACT REPORT

The work described in this report was done under ESA contract.

Responsibility for the contents resides in the author or organisation that prepared it.

Acronyms

BOA	Bottom Of Atmosphere
OSIP	Open Space Innovation Platform
TOA	Top Of Atmosphere
WP	Work Package

Table of Contents

1 Introduction

For decades humans have been designing, producing and mismanaging plastic items in ways and amounts to cause a global environmental issue. In order to better understand and then possibly mitigate the problem of plastic litter pollution, one of the first action is monitoring the problem.

Artificial Intelligence (AI) models can be trained to support such a monitoring action. There are not many applications of AI in this field, mainly due to the lack of available data (labelled or not) and the difficulty of acquiring a large number of validated data of litter accumulations at sea, large enough to be seen with remote sensing (in particular satellites).

As part of the Discovery Campaign on Remote Sensing of Plastic Marine Litter, Capgemini proposed an innovative way to use both its expertise in Artificial Intelligence and in Remote Sensing.

The goal of the project was to generate synthetic data sets of marine debris accumulations and to train on them AI models in view to automatically detect plastic litter accumulations in real EO images.

In order to achieve this goal, the project aimed at the following objectives:

1. To simulate plastic patches at different concentrations and shapes on real sea images coming from satellites such as Sentinel-2. The model is able to simulate the water leaving reflectance at different plastic concentrations and its effect on the spectral bands of Sentinel-2.

2. To detect marine plastic litter from simulated images from generated Machine Learning (ML) and/or Deep Learning (DL) models. The idea is to use ML/DL methods to automatically extract the specific information in spectral bands in order to detect marine plastic litter from simulated data.

3. To validate the simulated model with real plastic litter images through satellite images with marine plastic litter spots. The model, trained on simulated plastic litter images, will enable the validation with real EO images of plastic litter.

2 Methodology

Three data sets, represented in light blue colour in the following picture, corresponding to the reflectance spectra at the atmospheric levels of BOA (Bottom-Of-Atmosphere), TOA (Top-Of-Atmosphere) and Sensor Levels have been generated and used for training the AI algorithms. The following picture describes the simulation process.

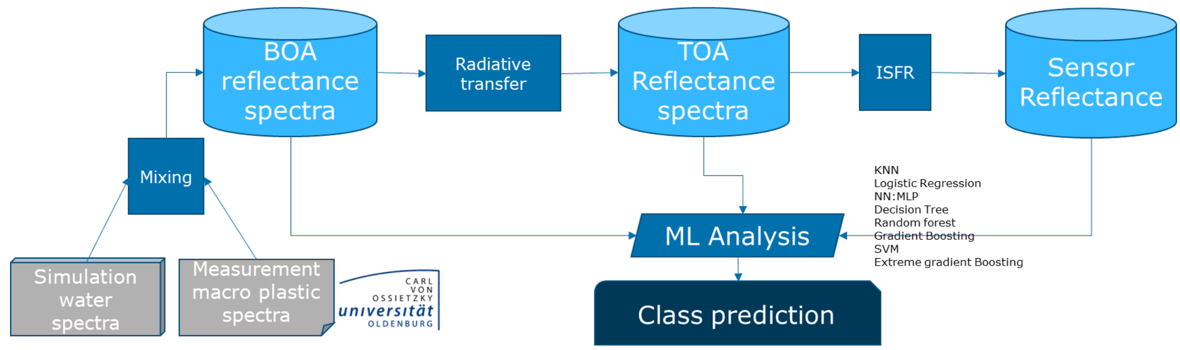

Figure 1 simulation workflow

Considering the difficulty to create a data set of real satellite images containing marine litter patches, due to the fact that validated images of this kind are at the moment very rare, we propose a simulation model performed in two steps:

Step 1: Simulation at “the pixel level”. Since the satellite sensors can measure several bands of the incident reflectance spectra, we first simulate at different levels of the atmosphere a mixture of reflectance spectra of sea water and surface macro-plastic litter. Therefore, the reflectance spectra are simulated by first simulating open ocean reflectance spectra taking into account different pigment concentrations, sun glints, and foam reflectances, and then linearly mixing the water reflectance with real macro-plastic reflectance spectra, for different plastic concentrations (with a minimum of 1% pixel area coverage).

Then the mixed reflectance spectra at the BOA level are transformed into the TOA level using a radiative transfer code. This step allows to add the impact of atmosphere (absorption and diffusion effect on the light signal). The TOA reflectance spectrum is the reflectance signal before being acquired by the sensor. To represent what the optic sensor captures, the spectra are convolved with the spectral response function of the sensor of interest to obtain the value for the corresponding bands. So the data set changes from continuous spectra to discrete spectra.

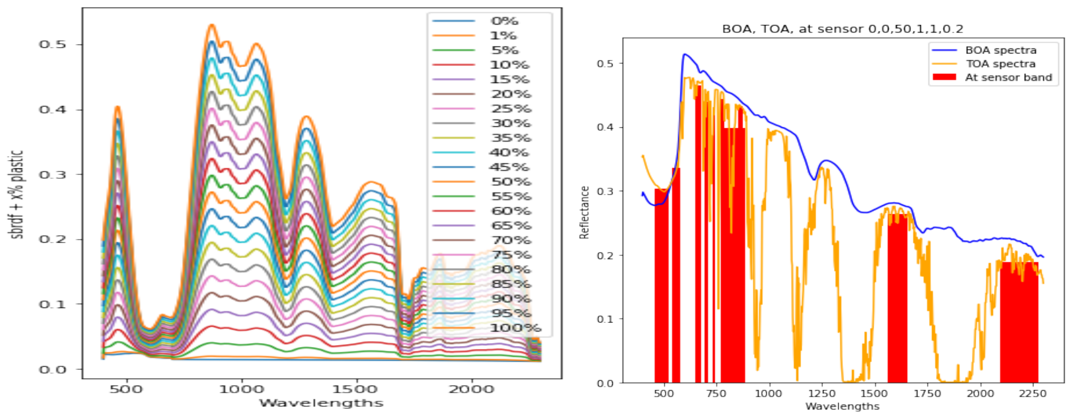

Figure 2: Spectra for different percentage of plastic mixing on the left and example of TOA, BOA and at sensor spectra on the right.

Then, we train AI models to differentiate spectra containing plastic from those containing only seawater on simulated data sets.

Step 2: Validation of AI models on real satellite images available. Three tiles of Sentinel-2 containing water and plastic targets are chosen. Pixels are labelled as containing plastic or not. Then, the AI models trained on the simulated data set are tested on these real Sentinel-2 images.

3 The results

For the validation purpose, the work was broken down into two phases:

1. Validation with simulated data sets of the trained AI models
2. Validation with real EO datasets of the most promising trained AI models

We have trained eight AI models to differentiate spectra containing plastic from those containing only seawater on these data sets. Only the AI models with more than 90% are selected. The results are shown in the table below.

accuracy	BOA		TOA		Sentinel2 10 bands		Sentinel2 4 bands	
ML model	Class without plastic	Class with plastic	Class without plastic	Class with plastic	Class without plastic	Class with plastic	Class without plastic	Class with plastic
KNN	94%	99%	36%	98%	91%	99%	**78%**	**98%**
Logistic Regression	**55%**	98%						
Neural Network	98%	99%	**17%**	99%				
Decision Tree	98%	99%	53%	98%	93%	99%	59%	98%
Random Forest	**99%**	**99%**	**69%**	**99%**	**97%**	**99%**	74%	98%
Gradient Boosting	**99%**	**99%**	32%	99%	80%	99%	34%	99%
Support Vector Machine	**55%**	98%						
Extreme Gradient Boosting	**99%**	**99%**	48%	99%	93%	99%	53%	99%

The main conclusion of this machine learning comparison is that the “Random Forest” is the best machine learning model for the data set at all levels and the accuracy of class with plastic is always higher due to the particular signature of plastic.

The results for Sentinel-2 bands are pretty good, it's mainly due to the fact that, at sensor level we have a discrete set of reflectances corresponding to the 10 bands (that are the features used). At BOA and TOA levels, as the spectra are continuous, we have defined specific features for the machine learning inputs.

The model tends not to detect plastic in concentrations of 1 to 5% in pixel area coverage, depending on the presence of sun glint and white caps automatically generated by the simulator. Further investigations are needed in order to better synthesise the influence of sun glint and white caps vs plastic concentrations.

Following the above analysis, we used only "Random Forest" model for the validation with real EO image for the final phase.

The validation of these models on available validated satellite images is the most challenging task of this project, due to the scarcity of validated EO images of plastic accumulations in the sea. We used the data set from Topouzelis et al. 2019[1].

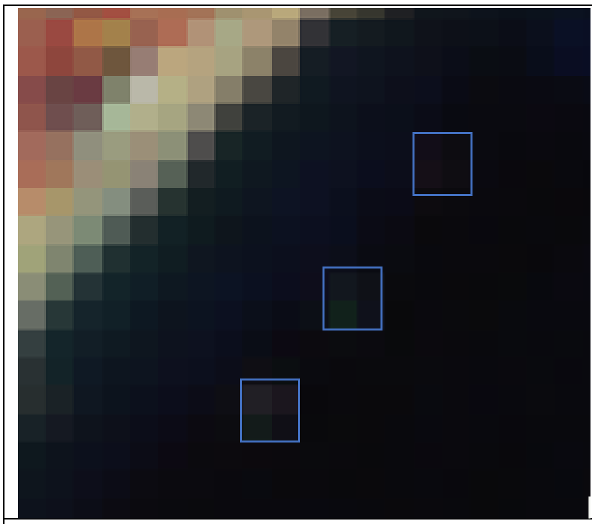	
Sentinel-2 image with 3 plastic patches (in the blue box)	Plastic prediction for random forest Sentinel2 10bands (grey for plastic, white for not plastic)

Figure 3: targets in the sea (left) and output of the AI algorithm (right)

Figure 3 shows that the three plastic patches are detected, but many pixels on land and seawater are also classified as plastic. Our model is trained on a simulated data set with an open ocean background and not coastal waters, so it turned to be not suitable for this validation. As the plastic bags targets have been placed on shallow waters partially above the sand and seagrass floor, the algorithm did not perform well.

The data set generated could be extended with new simulations adding more complexity and variety. Even if we focused on the macro plastic, the assumptions that we have chosen (the plastic is

[1] Topouzelis, K., Papakonstantinou, A., and Garaba, S. P. (2019) Detection of floating plastics from satellite and unmanned aerial systems (Plastic Litter Project 2018). International Journal of Applied Earth Observation and Geoinformation, v. 79, p. 175-183, doi:10.1016/j.jag.2019.03.011.

considered of zero thickness, opaque and floating on the water) can be improved, as for example to take into account the light reflection / transmission of the plastic in water. In addition, crossing the atmosphere to reach the sensor adds noise and in particular with the diffusion of radiation in some of the bands used for detection. We tested our models at different levels, simulation at surface level, in TOA and at the sensor. It would henceforth be judicious, in view to improve the results, to return to BOA, that is to say to apply an atmospheric correction on the data. Another improvement is to take into account the MTF (modulation transfer function) of Sentinel 2 and not only the spectral response of the instrument.

Particular and interesting emphasis should be on patches with low plastic pixel coverage where the signal to noise ratio would be critical for plastic detection. By injecting more complexity and "realism" and by harnessing the power of AI to extract plastic features in simulated patches containing not only plastic but also other materials like algae, driftwood, vegetation, foam and other floating matter found at sea, which can also have characteristic spectral signatures in the NIR-SWIR.