IMPROVED MARINE DEBRIS DETECTION IN SATELLITE IMAGERY WITH AN AUTOMATIC REFINEMENT OF COARSE HAND ANNOTATIONS

Marc Rußwurm, Dilge Gül & Devis Tuia
Environmental Computational Science and Earth Observation Laboratory (ECEO)
École Polytechnique Fédérale de Lausanne (EPFL)
Sion, Switzerland
correspondence to ma.russwurm@gmail.com

ABSTRACT

Plastic litter is a major environmental hazard that endangers human, animal, and plant health on the planet. A substantial portion of plastic pollutants is washed from rivers and beaches into the oceans and aggregates at the surface as marine debris before decomposing into microplastics and being digested by animals or sedimented on the sea floor. The marine debris is inherently challenging to annotate manually on satellite images, as the boundaries of floating objects are not sharp, and a specific mixture of water is always present at the pixel level. Hence, all available annotated marine debris datasets suffer from annotation errors. In this work, we present a label refinement algorithm for marine debris detection that improves upon rough hand annotations and considers the spectral characteristics of marine debris. We show quantitatively that a deep learning model trained with improved annotations achieves a higher classification accuracy on confirmed marine debris on two out of three datasets of confirmed plastic marine debris in Africa (in Ghana and South Africa). Thanks to the refinement module, we improve results for an environmentally important application that would benefit from further research attention to mitigate important associated challenges like label noise, domain shifts, and severe class imbalance.

1 INTRODUCTION

Microplastics are found across the entire planet. They have been shown to affect the growth of corals (Chapron et al., 2018) and were even detected in human stool (Schwabl et al., 2019). They enter the food chain in the oceans, where macro-plastics ($> 5\,\mathrm{mm}$ diameter) decompose into micro-plastics ($< 5\,\mathrm{mm}$ diameter) in the open water or during their transport in rivers (Van Emmerik et al., 2019; van Emmerik & Schwarz, 2020). Further, Van Dyck et al. (2016) demonstrated in beach surveys along the Accra-Temur Coastline in Ghana that beaches polluted by plastic debris are also a source of bacterial hazards, as shown by water samples taken simultaneously to the plastic surveys. Finally, many economic costs can also be associated with marine pollution, from clean-up expenses to loss of tourism revenue (Beaumont et al., 2019).

On open waters and under specific conditions, macroplastics can aggregate in elongated lines called *windrows*. These windrows are accumulations of surface debris. Ship-based collection along these features has proven highly effective, as demonstrated by Ruiz et al. (2020), who gathered 16.2 tons of floating marine litter in the Bay of Biscay, France, during a 68-working day campaign. These collection efforts are mainly scientific today, but rising economic demand for recycled ocean plastics may make a systematic collection of marine debris economically feasible. Magnier et al. (2019) have shown in a 2017 survey study in the Netherlands that sustainability-oriented consumers are interested and willing to pay a price premium for these products, and current campaigns raising awareness towards these issues are likely to increase the public sensitivity on this topic in future.

However, tracking marine debris in open waters is difficult (Cressey, 2016) and the lack of knowledge on the timely location of marine debris on the sea severely limits the efficiency of collection

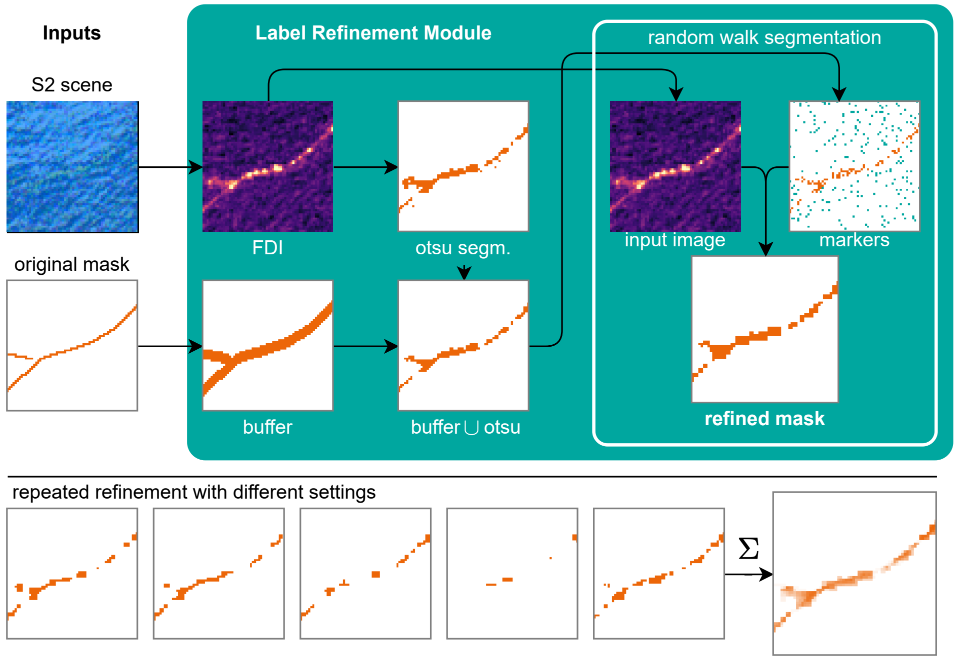


Figure 1: Label refinement module, producing a context-aware refined annotation mask from a Sentinel-2 image and a coarse original mask, hand-annotated.

efforts. Here, machine learning-based classification and detection models trained on open remote sensing data – as, for instance, the Copernicus Sentinel-2 satellite constellation – can support the collection efforts by providing up-to-date estimates of the location of marine debris at regular intervals.

To train these models on available Sentinel-2 imagery, comparatively large datasets of image scenes have been hand-annotated (Mifdal et al., 2021; Kikaki et al., 2022). However, the high manual effort in annotating this imagery and the lack of clear boundaries of marine debris make it challenging to produce a precise annotation map. For instance, Mifdal et al. (2021) annotated these elongated features exclusively as poly-lines of constant width rather than polygons. They then rasterized them to a single-pixel width ground reference map. While this often captures the general shape, it fails to capture variations in width that can sometimes be found in some large marine debris agglomerations.

In this work, we design a context-aware label refinement algorithm that inputs coarse hand annotations of marine debris and the Sentinel-2 image and provides an improved annotation mask that captures the true shape of marine debris in the dataset. It combines a classical semi-supervised computer vision processing pipeline with features based on our knowledge of the spectral signature of marine debris. We then use these refined labels as ground truth while training a conventional Unet++ segmentation model. We evaluate the performance of models trained with and without refined labels on two Sentinel-2 scenes of marine debris in Africa where marine debris has been confirmed to be present and on a dedicated Marine debris archive (MARIDA; Kikaki et al. (2022)) containing several annotated Sentinel-2 scenes. Our results indicate that adding an automatic refinement for these coarse annotations is generally beneficial, improving classification accuracy on two of three datasets.

2 Methodology

The label refinement module inputs an image patch of the multi-spectral Sentinel-2 (S2) image and the original hand annotations from the FloatingObjects dataset (Mifdal et al., 2021). Examples are shown in the top left of fig. 1. Given these inputs, it outputs a refined mask capturing the geometry of the marine debris visible in the Sentinel-2 scene (bottom right of fig. 1). Refining a single image-label pair is fast and we can repeat this process with different parameters to obtain multiple refined masks (bottom row of fig. 1), thus exploiting the uncertainty in the shape of marine debris during training the semantic segmentation model that learns from the refined masks.

The module itself consists of two stages: the first stage (left side of fig. 1) buffers the hand-annotated line to obtain a region of potential marine debris. It creates a first segmentation map of marine debris using the Otsu threshold (Otsu, 1979) on a "Floating Debris Index" (FDI) (Biermann et al., 2020) feature map. The FDI interpolates the spectral reflectance between the measured red-infrared edge (RE2: $782.8\,\mathrm{nm}$) and short-wave infrared (SWIR: $1613.7\,\mathrm{nm}$) signal and subtracts this interpolation from the measured near-infrared (NIR: $832.8\,\mathrm{nm}$) signal. In seawater, the interpolated NIR is typically close to the measured NIR, which leads to a low FDI, while marine debris has a higher response in near-infrared, leading to a high FDI. The buffer and segmentation are combined to obtain a preliminary area of marine debris near the original annotations. We use this preliminary area of marine debris to sample marker pixels, which are areas that confidently contain marine debris or non-debris (top right of fig. 1). These markers are the starting points of a computationally fast random walk segmentation algorithm (Grady, 2006). The markers are assumed to be accurately annotated, while the pixels between the markers are more uncertain. The random walker annotates these intermediate areas by an underlying anisotropic diffusion process that ensures that homogeneous areas are assigned to the same class. We apply this algorithm stochastically, meaning that one set of parameters (homogeneity criterion, buffer size, marker sampling frequency) of the random walker algorithm leads to one potential debris map, and running it with different parameter sets leads to a collection of potential refined masks.

We use the refinement module while training a UNET++ (Zhou et al., 2018) semantic segmentation model. We use a learning rate of 0.01 and weight decay 1×10^{-6} for 100 epochs. Before training, we generate target masks with the label refinement module with a buffer size of 0, 1, or 2 pixels, the β-parameter of the random walker (a penalization coefficient for the walker motion) of 1 or 10, and the marker density for marine debris of 5%, 25%, 50% or 75% (the density of *other* markers is fixed at 5%). Combined with the original mask, this yields 25 different target masks consistent with the hand annotations and the FDI image but of varying shapes and sizes.During training, we randomly choose one of these target masks as a kind of label augmentation. In our opinion, this reflects best the undefined borders of the marine debris that we aim to detect.

3 DATASET

We train on a combination of the two available marine debris datasets on Sentinel-2 data, namely the FloatingObjects dataset of Mifdal et al. (2021) and the Marine Debris Archive (MARIDA) collected by Kikaki et al. (2022). FloatingObjects contains 26 different globally distributed Sentinel-2 scenes. Overall, 3297 floating objects were annotated as lines when visually identified as *marine debris*. MARIDA contains 63 temporally overlapping Sentinel-2 scenes from 12 distinct regions. In total, 6672 polygons were annotated, of which 1882 are *marine debris* and 2447 *marine water*. The remaining 2343 polygons are annotated in one of 13 further classes, such as shallow water or ships, and were not added to the joint dataset. Additionally, we add images of ships without annotated marine debris as negative examples to our dataset. We use the S2Ships dataset of Ciocarlan & Stoian (2021), which segmented ships with Sentinel-2 imagery.

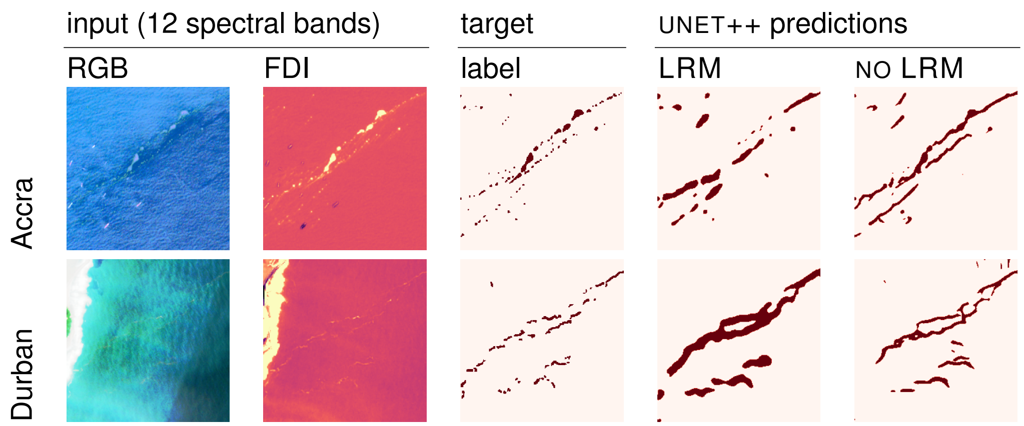
As a test set, we evaluate two Sentinel-2 scenes *Accra (2018-10-31)* and *Durban (2019-04-24)* that very likely contain plastics in the marine debris, as confirmed by social media and related studies (Van Dyck et al., 2016). In Accra, Ghana, beach surveys in 2013 showed that plastic materials made up the majority of 63.72% of marine debris washed onto evaluated beaches (Van Dyck et al., 2016). Further, visual inspections of the evaluation Sentinel-2 scene from 2018-10-31 and high-resolution imagery confirmed that beach waste deposits are subject to coastal erosion. The second evaluation scene from Durban, South Africa, 2019-04-24, shows marine debris being washed from the harbor. This image was taken shortly after a flood event that washed substantial amounts of plastic materials, as confirmed by the news and social media (Biermann et al., 2020). Finally, we use the test partition of the MARIDA archive as the third evaluation dataset.

4 EXPERIMENTS

Table 1a compares a UNET++ model trained on targets with label refinement module (LRM) and without the module (NO LRM) activated. All accuracy metrics slightly improved in the Durban

UNET++	Accra		Durban		MARIDA	
	LRM	NO LRM	LRM	NO LRM	LRM	NO LRM
ACCURACY	0.930 ± 0.016	$\mathbf{0.948} \pm 0.008$	$\mathbf{0.934} \pm 0.018$	0.905 ± 0.011	$\mathbf{0.867} \pm 0.005$	0.851 ± 0.006
F-SCORE	0.926 ± 0.018	$\mathbf{0.948} \pm 0.008$	$\mathbf{0.837} \pm 0.053$	0.776 ± 0.026	$\mathbf{0.749} \pm 0.009$	0.710 ± 0.015
AUROC	0.981 ± 0.006	$\mathbf{0.989} \pm 0.005$	$\mathbf{0.914} \pm 0.018$	0.886 ± 0.053	$\mathbf{0.746} \pm 0.021$	0.733 ± 0.006
JACCARD	0.862 ± 0.031	$\mathbf{0.900} \pm 0.014$	$\mathbf{0.722} \pm 0.048$	0.635 ± 0.034	$\mathbf{0.598} \pm 0.012$	0.551 ± 0.018
KAPPA	0.859 ± 0.031	$\mathbf{0.897} \pm 0.017$	$\mathbf{0.797} \pm 0.063$	0.717 ± 0.031	$\mathbf{0.661} \pm 0.012$	0.615 ± 0.017
improved?	no		yes		yes	

(a) Quantitative comparison



(b) Qualitative predictions on each $2.56\,\text{km}$ by $2.56\,\text{km}$ scenes from the Accra and Durban.

Table 1: Comparison of UNET++ models with and without the label refinement module activated.

scene and on the MARIDA test set. However, the model without the refinement module achieved slightly better metrics in the Accra scene. In table 1b, we show two qualitative examples from the Accra and Durban scenes. Both models capture the general shape of marine debris visible in the FDI representation of the Sentinel-2 scene. However, the fine-grained individual patches of marine debris are not always accurately captured since both models smooth the prediction mask. The model with refinement predicts larger patches of marine debris than without refinement. This is an effect of the label refinement module always enlarging the original 1-pixel line annotations into patches and, hence, biases the model to predict larger patches on average.

In summary, we can see a positive effect of the label refinement module on the MARIDA scenes and Durban. However, this effect is smaller than hoped, as only a small improvement was measured compared to the model without LRM.

5 Discussion and Conclusion

Detecting marine debris accurately and at a large scale from readily available satellite data is key to efficient clean-up and collection efforts of plastic litter in open waters. These efforts are becoming increasingly important fields with further increasing levels of pollution of our waters and may become economically viable due to increased sensitivity for sustainable economics and increased demand for recycled plastics, for instance, in clothing.

We tested a label refinement module specifically designed for the characteristics of marine debris. Training an off-the-shelf deep learning model with refined annotations from this module leads to better results on two of three evaluation datasets. The results in this work highlight the benefit of designing components of a machine learning model specifically for a given problem. The fact that we could not improve the detection across all three datasets highlights the difficult and diverse nature of marine debris in open waters that vary in composition and appearance from region to region. Further canonical research is necessary to build reliable marine debris detectors and to identify sources of ocean pollution with available data.

REFERENCES

Nicola J Beaumont, Margrethe Aanesen, Melanie C Austen, Tobias Börger, James R Clark, Matthew Cole, Tara Hooper, Penelope K Lindeque, Christine Pascoe, and Kayleigh J Wyles. Global ecological, social and economic impacts of marine plastic. *Marine pollution bulletin*, 142:189–195, 2019.

Lauren Biermann, Daniel Clewley, Victor Martinez-Vicente, and Konstantinos Topouzelis. Finding plastic patches in coastal waters using optical satellite data. *Scientific reports*, 10(1):1–10, 2020.

Leila Chapron, Erwan Peru, Adam J. Engler, Jean-François Ghiglione, Anne-Leila Meistertzheim, Audrey M Pruski, Autun Purser, Gilles Vétion, Pierre E Galand, and Franck Lartaud. Macro-and microplastics affect cold-water corals growth, feeding and behaviour. *Scientific reports*, 8(1):1–8, 2018.

Alina Ciocarlan and Andrei Stoian. Ship detection in sentinel 2 multi-spectral images with self-supervised learning. *Remote Sensing*, 13(21):4255, 2021.

Daniel Cressey. The plastic ocean. *Nature*, 536(7616):263–265, 2016.

Leo Grady. Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1768–1783, 2006.

Katerina Kikaki, Ioannis Kakogeorgiou, Paraskevi Mikeli, Dionysios E Raitsos, and Konstantinos Karantzalos. Marida: A benchmark for marine debris detection from sentinel-2 remote sensing data. *PloS one*, 17(1):e0262247, 2022.

Lise Magnier, Ruth Mugge, and Jan Schoormans. Turning ocean garbage into products–consumers' evaluations of products made of recycled ocean plastic. *Journal of cleaner production*, 215:84–98, 2019.

Jamila Mifdal, Nicolas Longépé, and Marc Rußwurm. Towards detecting floating objects on a global scale with learned spatial features using sentinel 2. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-3-2021:285–293, 2021. doi: 10.5194/isprs-annals-V-3-2021-285-2021.

Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

Irene Ruiz, Oihane C Basurko, Anna Rubio, Matthias Delpey, Igor Granado, Amandine Declerck, Julien Mader, and Andrés Cózar. Litter windrows in the south-east coast of the Bay of Biscay: an ocean process enabling effective active fishing for litter. *Frontiers in marine science*, 7:308, 2020.

Philipp Schwabl, Sebastian Köppel, Philipp Königshofer, Theresa Bucsics, Michael Trauner, Thomas Reiberger, and Bettina Liebmann. Detection of various microplastics in human stool: a prospective case series. *Annals of Internal Medicine*, 171(7):453–457, 2019.

Irene P Van Dyck, Francis KE Nunoo, and Elaine T Lawson. An empirical assessment of marine debris, seawater quality and littering in Ghana. *Journal of Geoscience and Environment Protection*, 4(5):21–36, 2016.

Tim van Emmerik and Anna Schwarz. Plastic debris in rivers. *Wiley Interdisciplinary Reviews: Water*, 7(1):e1398, 2020.

Tim Van Emmerik, Romain Tramoy, Caroline Van Calcar, Soline Alligant, Robin Treilles, Bruno Tassin, and Johnny Gasperi. Seine plastic debris transport tenfolded during increased river discharge. *Frontiers in Marine Science*, 6:642, 2019.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested U-Net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 3–11. Springer, 2018.