



Dataset Documentation

Dataset Name: MARIDA: Marine Debris Archive Dataset

Description

Marine Debris Archive (MARIDA) is a marine debris-oriented dataset on Sentinel-2 satellite images. It also includes various sea features (clear & turbid water, waves, etc.) and floating materials (*Sargassum* macroalgae, ships, natural organic material, etc) that co-exist. MARIDA is primarily focused on the weakly supervised pixel-level semantic segmentation task.

Citation

Kikaki K, Kakogeorgiou I, Mikeli P, Raitsos DE, Karantzalos K (2022) MARIDA: A benchmark for Marine Debris detection from Sentinel-2 remote sensing data. PLoS ONE 17(1): e0262247. https://doi.org/10.1371/journal.pone.0262247

License

CC-BY-4.0

Creator(s) National Technical University of Athens

Contact Katerina Kikaki: akikakh@central.ntua.gr

Tutorial

Quick Start Guide for MARIDA (Marine Debris Archive) on Github

Publications

MARIDA: A benchmark for Marine Debris detection from Sentinel-2 remote sensing data

Location and boundaries

Overall Location Method

- Ground collection only
- Ground collection with boundary drawn using imagery
- Ground collection with spatial buffer added
- Boundary drawn from imagery
- Other _____
- Unknown

GeoLocation Device





- Industrial grade GPS (List model) _____
- Retail grade GPS
- Mobile Phone GPS
- N/A
- Unknown

Ground Boundary Method (Details explained in Appendix A)

- Live/Continuous point capture of walk-around
- Manual point capture of walk-around
- Manual point capture of polygon boundaries (not whole field)
- Manual point capture f4or later image annotation
- Manual point capture for spatial buffer within field
- Manual point capture while looking at but not in field, with heading recorded
- Other _____
- Unknown

Imagery used

Sensor: Sentinel-2

Date(s): 2015-11-29 to 2021-01-23

List Sentinel-2 tiles used in Appendix B

Imagery Annotation methods

- Boundaries drawn based on a single ground point captured
- Boundaries drawn/edited based on multiple ground points captured
- Buffer validated from ground point captured
- Boundary drawn without ground reference data (Include description of methods in Appendix C)
- Pixels annotated without ground reference data (Include description of methods in Appendix C)
- Unknown

Boundary inclusion

- Captured polygon includes the entire field/area
- Captured polygon includes only a sample of the field/area

Classification

Classification Type

- Land cover
- Crop type
- ☑ Other: Sea surface: Marine Debris and various sea features and objects that co-exist





Classes/fields used

Describe in Appendix D

Ground Referenced Classification

- Observation (Describe methods of determination in Appendix E)
- Survey/interview with land holder (Describe methods in Appendix E)
- Other (Describe methods in Appendix E)

Image Referenced Classification

Describe methods used in Appendix C

Appendix A: Describe the method of geographic ground data collection

Most of the ground data was collected by citizen scientists, photographers, and NGOs on vessel expeditions or clean-up activities in the Caribbean Sea (Guatemala, Honduras, Haiti, S. Domingo) and Asia (Indonesia, Vietnam, Philippines). Additionally, some of the annotated data was verified by past studies (Kikaki et al., 2020, Biermann et al., 2020) who collected marine debris events in the Bay islands of Honduras, and other regions such as Scotland and S. Africa.

Appendix B: Collected Marine Debris Ground-Truth Data by Source

Continent/ Country	S2 Tile	Source	Date	Location (WGS'84)
C. America/ Guatemala	16PCC	Citizen Scientist	18/9/20	15.836206° N, 88.022087° W
C. America/ Guatemala	16PCC	Photographer	16/6/18	15.827222° N, 88.047500° W
C. America/ Guatemala	16PCC	Kikaki et al. (2020)	4/9/19	14.9827° N, 89.5442° W
C. America/ Honduras	16PDC	Citizen Scientist	18/9/20	16.1490° N, 87.6282° W
C. America/ Honduras	16PEC, 16QED	Citizen Scientist	23/9/20	16.042194° N, 86.432081° W
C. America/ Honduras	16PEC	Kikaki et al. (2020)	29/11/15	16.0667° N, 86.3965° W
N. America/ S. Domingo	19QDA	Media	13/7/18	18.467723° N, 69.886808° W
N. America/ Haiti	18QWF/ QYF/ QYG	4ocean Clean-Ups	20/3/20	-
N. America/ Haiti	18QWF/ QYF/ QYG	4ocean Clean-Ups	5/1/21	-
N. America/ Haiti	18QWF/ QYF/ QYG	4ocean Clean-Ups	9/12/20	-
N. America/ Haiti	18QWF/ QYF/ QYG	4ocean Clean-Ups	15/12/20	-
Asia/ Indonesia	50LLR	Social Media	4/3/18	8.715828° S, 115.446799° E
Asia/ Vietnam	48PZC	Social Media	23/11/19	15.994762° N, 108.27417° E
Asia/ Philippines	51PTS	Social Media	18/5/19	-
Asia/ Philippines	51PTS	Social Media	16/7/16	-
Europe/ Scotland	30VWH	Biermann et al. (2020)	20/4/18	-
Africa/ South Africa	36JUN	Biermann et al. (2020)	24/4/19	-
Asia/ South Korea	52SDD	Jang et al. (2014)	-	-
Asia/ Indonesia	48MXU/ MYU	Cordova & Nurhati (2019)	-	-
Asia/ China	51RVQ	Zhao et al. (2019)	-	-





Appendix C: Describe how boundaries and classes were determined without ground reference data

First ground-truth events were collected, reported cases were then identified in S2 satellite data, and verified with very high-resolution satellite data. Well-established spectral signatures were also calculated to improve the annotation step. The annotators digitized Marine Debris based on ground-truth events, and employing domain knowledge about its spectral behavior and its accumulation patterns (i.e., fronts, marine litter windrows). A laborious and intensive image interpretation and manual assessment of each pixel were performed for all selected images leading to Marine Debris annotations at pixel level. In addition, diverse floating objects, sea state features, water types and clouds were annotated based on image interpretation and established spectral and spatial patterns. Wind data were utilized to examine the possibility of whitecaps, which may appear similar to plastics in human eye. An inter-annotator agreement protocol was also established to merge the annotated data and aggregate the confidence levels derived from the annotators. Finally, a detailed spectral and statistical analysis was performed and assessed to refine the data and improve as much as possible the quality of MARIDA dataset. The spectral signatures analysis and the detailed overview of MARIDA can be found here: https://marine-debris.github.io/. Please refer to Appendix D for a complete list of the top-level classes used to distinguish marine debris from ships, macroalgae, waves, water types, clouds, cloud shadows, etc.

Appendix D: List all top-level classes or the classification guidance used

- 1: 'Marine Debris'
- 2: 'Dense Sargassum'
- 3: 'Sparse Sargassum'
- 4: 'Natural Organic Material'
- 5: 'Ship'
- 6: 'Clouds'
- 7: 'Marine Water'
- 8: 'Sediment-Laden Water'
- 9: 'Foam'
- 10: 'Turbid Water'
- 11: 'Shallow Water'
- 12: 'Waves'
- 13: 'Cloud Shadows'



14: 'Wakes'

15: 'Mixed Water'

