Open Ocean Data Collection for AI and ML bubu

An Open Ocean Data Collection for AI and ML

A curated collection of open datasets and datasources related to the ocean, marine biology and climate change.

Contributing

We welcome contributions to this guide. Please read a `CONTRIBUTING.md` on how to proceed.

Genomic sequence datasets

Plankton images datasets

Ecotaxa

Ecotaxa Picheral et al. (2017) that images contains over 188 million "objects" caputured by different microscopes. Part of the dataset has been manually annotated. They also have an ML-based image classifier.

WHOI-Plankton: 3.5 million human annotated plankton imgs

Annotated Plankton Images - Data Set for Developing and Evaluating Classification Methods

The data set available here comprises > 3.5 million images of microscopic marine plankton, organized according to category labels provided by researchers at the Woods Hole Oceanographic Institution (WHOI). The images are currently placed into one of 103 categories.

The annotated image set can be downloaded at the following citable URI: http://dx.doi.org/10.1575/1912/7341.

The images are part of a much larger data set (>700 million images) collected in situ by automated submersible imaging-in-flow cytometry with an instrument called Imaging FlowCytobot (IFCB) (Olson & Sosik, 2007; Sosik & Olson, 2007) at the Martha's Vineyard Coastal Observatory (MVCO) starting in 2006 and continuing to the present. Near real time image data and the complete archive are accessible for browse and download at the IFCB Data Dashboard.

Satelite imaging datasets

google earth, data observatory, sentinel, copernicus, etc.}

Papers datasets

(Lima-Mendez et al., 2015) sdfasdfasdf

References

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J. C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic, S., Berline, L., Bontempi, G., Cabello, A. M., Coppola, L., Cornejo-Castillo, F. M., ... Raes, J. (2015). Ocean plankton: Determinants of community structure in the global plankton interactome. *Science*, *348*(6237), 1262073. https://doi.org/10.1126/science.1262073

Olson, R. J., & Sosik, H. M. (2007). A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot. *Limnology and Oceanography: Methods*, *5*(6), 195–203. https://doi.org/10.4319/lom.2007.5.195

Picheral, M., Colin, S., & J.-O., I. (2017). *EcoTaxa, a tool for the taxonomic classification of images*. http://ecotaxa.obs-vlfr.fr

Sosik, H. M., & Olson, R. J. (2007). Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. *Limnology and Oceanography: Methods*, *5*(6), 204–216. https://doi.org/10.4319/lom.2007.5.204