AI in marine sciences: an open-access integrated environment for automated classification of phytoplankton images

Lagaisse Rune, Wout Decrop, Klaas Deneudt

Marine Observation Centre, Research department, Flanders Marine Institute (VLIZ), InnovOcean Campus, Jacobsenstraat 1, 8400 Oostende, Belgium
E-mail: rune.lagaisse@vliz.be

Phytoplankton, the single cell algae at the basis of marine food webs and an important indicator of ecosystem health, is continuously being monitored at a number of stations in the Belgian Part of the North Sea under the LifeWatch research infrastructure. To process monitoring samples in a fast and automated manner, we make use of automated imaging techniques like FlowCam. By aligning particles in the sample in a continuous fluid stream and capturing each particle in a picture as it passes a camera, this device can produce an image library of a sample in under 30 minutes. While FlowCam has significantly sped up time spend in the lab, it delivers about 350 000 images on a yearly basis and calls for an automated approach to handle high data loads. To speed up taxonomist's job of labeling all these images manually, we build semi- automated data pipelines and implemented machine-learning algorithms, more specifically Convolutional Neural Networks, to classify the images. Over the years this combination of automated imaging and machine learning has helped us built a set of over 2,2 million annotated FlowCam images and trained classifiers fine-tuned by taxonomists correcting wrong model predictions. This dataset and the trained classifiers have proven to be a huge benefit in our marine monitoring, and we wanted to share this asset with other researchers.

Under the Horizon Europe iMagine project, we aim to not only publish the image set and classifiers, but to build a user-friendly module where users can both predict FlowCam images using pretrained models and train classifiers on their own image input. The iMagine platform hosting this module offers an integrated environment with all source code as well as a graphical user interface for users with less coding experience. Computing resources for the services are also available to the user though the platform. The FlowCam module further provides tools for post-hoc analysis of model performance and code for image transformation and augmentation to deal with different image resolutions and class imbalances in training sets. More information on the project and the FlowCam service can be found via https:// www.imagine-ai.eu. In the next coming years, we hope to facilitate many marine researchers in the application of automated classification of phytoplankton imaging data. We actively encourage researchers and monitoring programs to make use of the FlowCam service and the iMagine platform to contribute to more efficient biomonitoring.

Keywords
AI; ML; CNN; FlowCam; Phytoplankton; IMagine