# From pixels to patterns: high-throughput image classification and morphometry through a new PI-10 imaging pipeline

Jonas Mortelmans[a], Wout Decrop[a], Hanneloor Heyndericx[a], André Cattrijsse[a], Mark Depaepe[b], Lodewijk Van Walraven[c, d], James Scott[e], Dick van Oevelen[f], Klaas Deneudt[a], Carlota Muñiz[a]

[a] Flanders Marine Institute (VLIZ) InnovOcean Campus, Jacobsenstraat 1, 8400 Ostend, Belgium.
[b] VLOOT dab, Sir Winston Churchillkaai 2, 8400 Oostende, Belgium.
[c] Wageningen University and Research (WUR), Wageningen Marine Research, P.O. Box 57, 1780 AB Den Helder, the Netherlands
[d] Rijkswaterstaat (RWS), Postbus 2232, 3500 GE Utrecht. The Netherlands.
[e] Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Pakefield Road, Lowestoft NR33 0RP, the United Kingdom.
[f] Department of Estuarine and Delta Systems, Royal Netherlands Institute for Sea Research (NIOZ), PO Box 140, 4400 AC, Yerseke, the Netherlands

Corresponding author: Jonas Mortelmans; jonas.mortelmans@vliz.be

## Introduction

Plankton are ubiquitous and their importance to marine food webs are thoroughly documented. Many plankters are especially vulnerable to climate-driven changes: their phenology, range, and body size change with climate having consequences for the biological carbon pump and higher trophic interactions. Plankton size structure, from individual body length to community-level size spectra, affects trophic efficiency and predator-prey dynamics. Finally, being a chain in the marine food web, plankters are essential for ocean's health and resilience.

The Belgian part of the North Sea (BPNS), a small but dynamic marine coastal area, is hosting a rich planktonic community shaped by strong seasonal cycles, riverine inputs, eutrophication, and climate variability. Historically, plankton monitoring in the BPNS has relied on net sampling and manual microscopy, labour-intensive methods, relatively limited in resolution and prone to observer bias. Recent decades have seen a rapid shift in marine observation capabilities, particularly through the development of automated, high-resolution imaging technologies able to yield size and biomass estimations. A particularly interesting new and emerging technology is the Plankton Imager 10 (Pi-10, developed by the Centre for Environment Fisheries and Aquaculture Science (CEFAS) and Plankton Analytics), a line-scan imaging system. The Pi-10 is a standalone, automated imaging instrument designed for continuous, in situ monitoring of plankton and particles in a flow-through system processing around 1m³ per 29 minutes.

In this study, we present a modular and reproducible workflow for processing images acquired with the PI-10. The pipeline integrates convolutional neural network (CNN)-based classification, morphological feature extraction, and standardized data publication via Darwin Core Archives. Special attention is given to the extraction of size-related metrics such as equivalent spherical diameter, eccentricity, and major/minor axis lengths, allowing trait-based ecological interpretation. Its high-frequency, high-volume data collection makes it an ideal tool for studying fine-scale ecological processes and for contributing to the development of digital ocean twins.

**Methods**

The Pi-10 (fig. 1) captures high-resolution images of particles as they pass through a flow cell. A downward-facing line-scan camera records up to 70,000 lines per second with 2048 × 10 µm resolution. These lines are concatenated based on the water flow and yields regions of interest (ROIs). Images are transmitted via fiber optic cable to the onboard Pi-PC for real-time processing and storage. The system can detect up to 100,000 regions of interest (ROIs) per minute based on a user dependant size cutoff level (for zooplankton ecology, typically set at a threshold of 200µm; but can be as low as 20µm). The PI-10 is mounted within a temperature- and humidity-controlled metal frame and part of the ships' underway data acquisition system, easily linked to metadata/data streams. On RV Simon Stevin, water intake is at 0.6 m depth—ideal for surface monitoring.

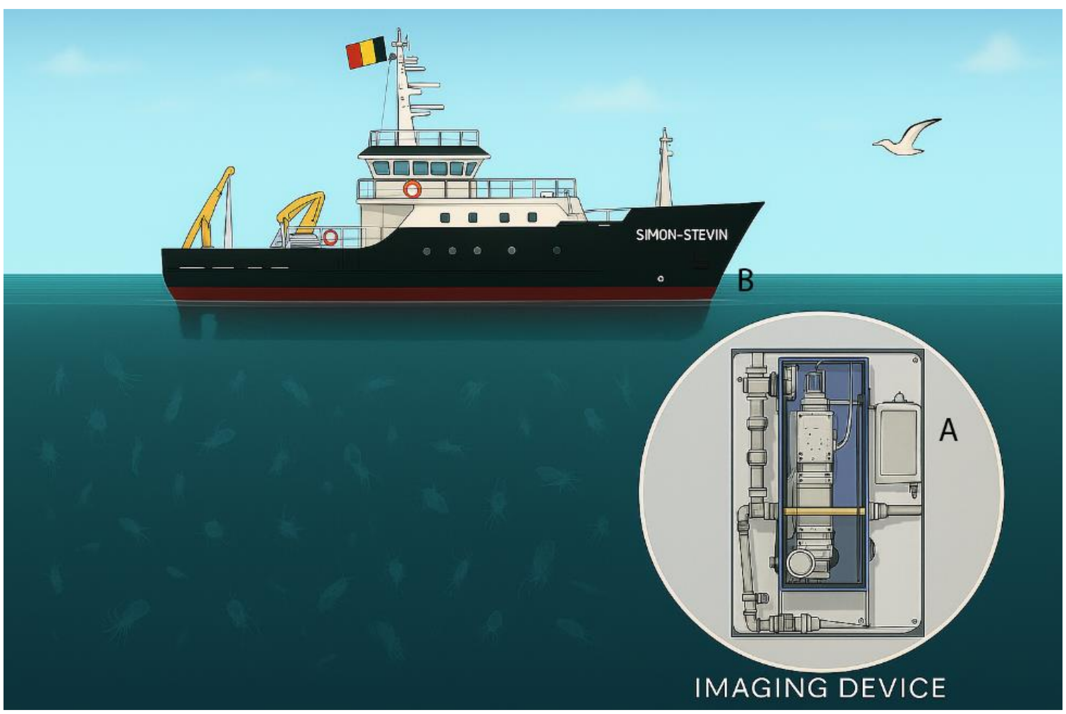
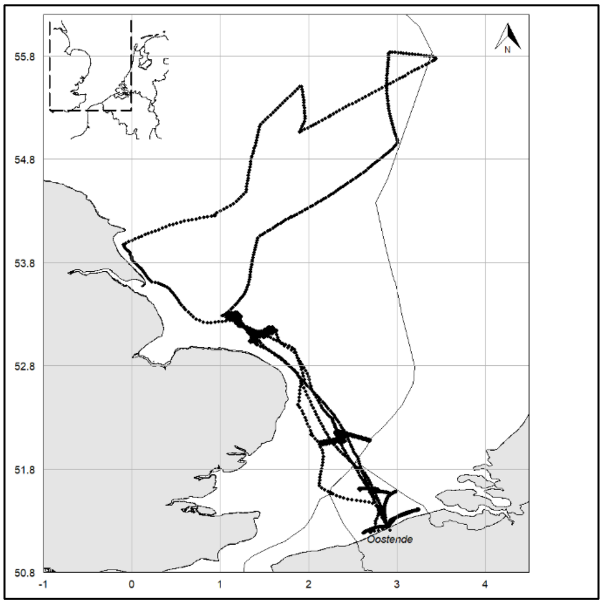



*Fig. 1. Illustration of the RV Simon Stevin (B) and PI-10 sensor (A)*

**Results**

Data was collected during 9 cruises between March 2025 and July 2025 aboard the RV Simon Stevin in the southern North Sea (Fig. 2) and yielded 1.2 billion images of which 60 million are saved and 1 million are validated manually for testing and model accuracy estimates (example seen in Fig. 3).



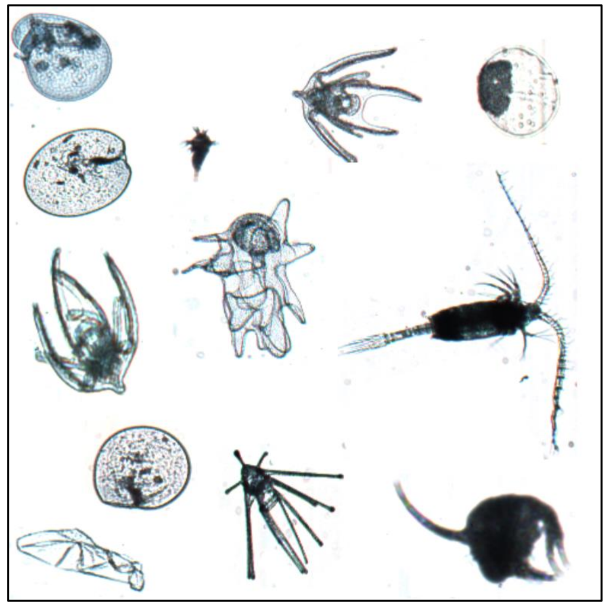
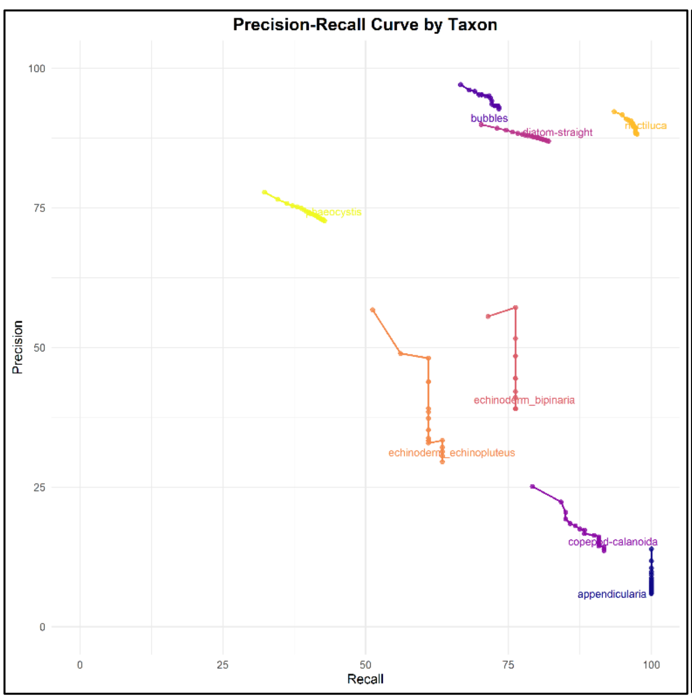



*Fig 2. Transects on which the PI-10 was deployed. Fig 3. Representation of most common taxa visualised by the PI-10*

The here published PI-10 image processing pipeline integrates a modular script designed to handle data ingestion, classification, validation, and standardised output, with all code openly published (Mortelmans et al., 2025). Upon acquisition, raw .tar archives are securely stored on a VLIZ server. The pipeline begins with a preview stage in which 200 randomly selected .tif images are classified. Archives containing >30% predicted "bubbles" are flagged as low quality and excluded, saving processing time by leveraging the model's high precision for this class. In total, ten output types are subsequently generated, including: (1)_background, containing the background image of a .tar file, essential for quick assessment of the metadata; (2) _gpstag, containing a timestamp and GPS coordinates for each ROI in the .tar; (3) _hitsmisses, to keep track of the fraction imaged; it counts the amount of hits and misses, in each minute of the .tar file; (4) _image_properties; an extension to store all metrics on a certain ROI (e.g., eccentricity, ESD, major, minor, ...); (6) _predictions_relative, a file containing the result of the CNN-algorithm; (7) _topspecies, an extension to _predictions_relative, to store the best fitting class to each ROI, according to a user defined threshold; (8) _validated, the result of a validations; (9) _merged, combined the different outputs; (10) _dwca, compiles the validated dataset into a

Darwin Core Archive (DwC-A), structured for interoperability with biodiversity data platforms such as EurOBIS. The DwC-A contains an Event Core (sampling metadata), Occurrence Core (species records linked to WoRMS taxonomy), and an eMoF extension for environmental parameters like temperature and salinity; but also include total estimated biomass. Final archives are published in the Marine Data Archive (MDA) with a DOI, ensuring traceability and open access. All datasets are discoverable through IMIS, VLIZ's ISO-19115-compliant metadata catalogue.

The PI-10 zooplankton image classifier uses a Convolutional Neural Network (CNN) based on the FlowCam phytoplankton model developed in the iMagine project (Decrop *et al.*, 2025), here trained on 350,000 annotated PI-10 images provided by Van Walraven *et al.* (2025). The PI-10 classifier adopts this architecture, reusing the AI platform, OSCAR services, and FAIR metadata standards. The classifier's output performance varies across taxonomic groups (Fig 4, 5). Certain classes, such as *Noctiluca*, bubbles, and straight diatoms, are identified with high accuracy and precision, exhibiting minimal false positives even at low confidence thresholds. In contrast, classes like Appendicularia and Calanoida show high recall but low precision, with frequent misclassification of other taxa into these categories. Finally, several taxa, such as *Phaeocystis*, are poorly resolved by the model (Fig. 4, 5).



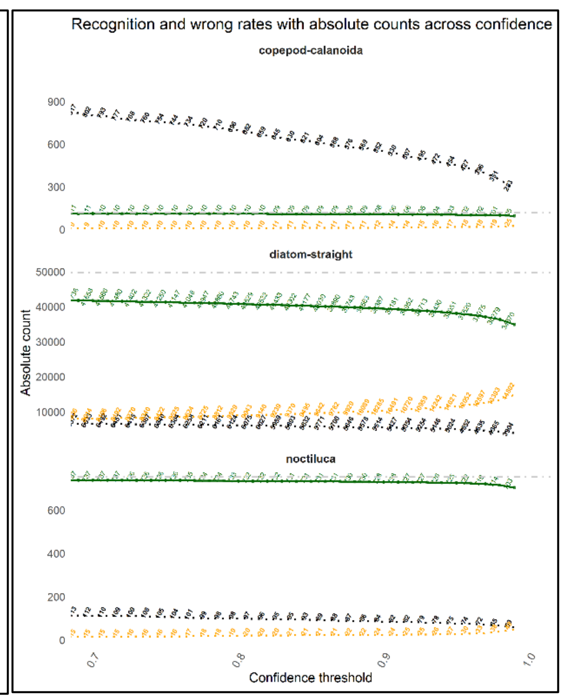




*Fig 4. Precision-recall curves for main biological taxa. Fig 5. Recovery rates of correct and false predictions (three taxa only). Green: correctly identified ROIs. Yellow: wrongly identified ROIs. Black: False predictions.*

## Discussion

A key strength of the PI-10 pipeline lies not only in its taxonomic classification capabilities but also in its integration of standardized size measurements for every detected particle. This goes beyond the focus of many image-based biodiversity tools, which often concentrate solely on species detection. Size and morphological traits offer critical ecological context and play an essential role in functional ecology and ecosystem modeling (e.g., Benedetti et al., 2018). Quantifying size structure is particularly important because it directly reflects community dynamics, which influence trophic interactions, grazing pressure, and biogeochemical fluxes (Hirai et al., 2020; Lombard et al., 2019). Moreover, accurate size data enables biomass estimation and functional trait analyses that link organism size to physiological processes, feeding strategies, and motility patterns. Finally, size-based indicators derived from imaging systems are increasingly used in marine ecosystem monitoring and biodiversity assessments (Benedetti et al., 2021). It is obvious, the size-measurements of in situ particles here provided are essential.

The performance of the CNN classifier across specific taxa is variable. For instance, distinguishing between morphologically similar or irregular taxa remains challenging. This reflects a broader trade-off between automation and expert validation effort; especially seen in the result of the currently published classifiers. For taxa, alike *Noctiluca* or straight diatoms, it can easily be argued to skip human validation entirely (see fig. 4 and 5). The model holds high accuracy and precision, exhibiting minimal false positives even at low confidence thresholds. For other taxa, Appendicularia or Calanoida, are correctly identified even at low confidence thresholds; but hold many false positives. Finally, taxa like *Phaeocystis*, the model has both accuracy and precision low. For these two groups of taxa, manual validation remains essential. Despite automation significantly reducing manual annotation workload (Luo et al., 2018), the risk of misclassification, especially for key ecological indicators, necessitates continued expert validation.

To further advance the PI-10 pipeline, model improvement efforts should focus on integrating temporal context, such as seasonal variability in plankton communities. This could involve seasonal model retraining or transfer learning approaches. Moreover, there is strong potential for linking image-derived traits to ecosystem indicators (e.g., biodiversity indices, bloom risk alerts, or ecosystem health assessments) as highlighted in recent studies (Lombard et al., 2019; Ohman et al., 2019; Irisson et al., 2022). A promising direction involves real-time or in situ deployment, with onboard image classification modules enabling near-instant data feedback for adaptive sampling or early warning systems (Irisson et al., 2022; Pitois et al., 2025). Such technological advances can improve responsiveness in oceanographic research and monitoring, while also supporting integration with satellite and modeling frameworks (Ohman et al., 2019). The automation of ecological interpretation from plankton images remains an emerging but essential frontier in this field (Lombard et al., 2019).

**Conclusion**

This work assesses the practicalities and on-board implementation of the Pi-10 on the RV Simon Stevin, exploring its integration with the ship's existing underway system, while generating the first collection of high-resolution PI-10 imagery and annotations for the Belgian part of the North Sea. A fully open-access data processing pipeline and classification framework is presented combining geotagged imaging, deep learning (CNN-based), image-based metrics and a multi-label taxonomic system. This initiative represents a step toward automated, real-time data collection and downstream ecological assessments in coastal systems - with its applicability well beyond the BPNS. By quantifying plankton, and suspended particles as a whole, size at high resolution critical insights into community structure are detected - making size not just a descriptor, but a key ecological indicator of change.

**References**

Barth, A. J., Pinckney, J. L., Krask, J., Smith, E., & Stone, J. (2025). Storm in a bottle: An experimental investigation of how extreme precipitation impacts phytoplankton communities. *bioRxiv*. https://doi.org/10.1101/2025.05.20.655176

Benedetti, F., Gasparini, S., & Irisson, J.-O. (2018). The seasonal and inter-annual fluctuations of plankton abundance and community structure in the Northwestern Mediterranean Sea. *Progress in Oceanography*, 162, 1–12.

Benedetti, F., Righetti, D., & Irisson, J.-O. (2021). Major restructuring of marine plankton assemblages under global warming. *Nature Communications*, 12, 5229.

Decrop, W., Deneudt, K., Heredia, I., Calatrava Arroyo, A., Lagaisse, R., & Muñiz, C. (in preparation). *Design and implementation of an automated image classification workflow for phytoplankton monitoring*.

Hirai, J., Tsuda, A., & Suzuki, K. (2020). Global gradients in species richness of marine plankton functional groups. *Journal of Plankton Research*, 45(6), 832–845

Irisson, J.-O., Kraft, K., & Velhonoja, O. (2022). Towards operational phytoplankton recognition with automated high-throughput microscopy and machine learning. *Frontiers in Marine Science*, 9, 867695.Lagaisse et al., 2025

Lombard, F., Boss, E., Waite, A. M., Uitz, J., Stemmann, L., Sosik, H. M., et al. (2019). Globally consistent quantitative observations of planktonic ecosystems. Front. Mar. Sci. 6. doi: 10.3389/fmars.2019.00196

Luo, T., Suikkanen, S., & Tamminen, T. (2018). Automated plankton image classification using deep learning. *Frontiers in Marine Science*, 5, 370

Mortelmans, J., Decrop, W., Heynderickx, H., Cattrijsse, A., Depaepe, M., Van Walraven, L., Scott, J., van Oevelen, D., Deneudt, K., & Muñiz, C. (n.d.). *From pixels to patterns: High-throughput image classification and morphometry through a new PI-10 imaging pipeline*.

Ohman, M. D., Davis, R. E., Sherman, J. T., Grindley, K. R., Whitmore, B. M., Nickels, C. F., & Ellen, J. S. (2018). Zooglider: An autonomous vehicle for optical and acoustic sensing of zooplankton. Limnology and Oceanography: Methods, 17(1), 69–86. https://doi.org/10.1002/lom3.10301

Pitois, S.G., Blackwell R.E., Close H., Eftekhari N., Giering, S.L.C, Masoudi M., Payne E., Ribeiro J, Scot, J. (2025). RAPID: real-time automated plankton identification dashboard using Edge AI at sea. Frontiers in Marine Science 11. https://doi.org/10.3389/fmars.2024.1513463

Van Walraven, L., Scott, J., Pitois, S., CEFAS team, Hoekendijk, J., Hovenkamp, P., Jak, R. and van Oevelen, D., 2025. A large (50k+) and FAIR training set for the Plankton Imager (Pi-10) for the Greater North Sea and NW Atlantic, based on a flexible classification protocol.