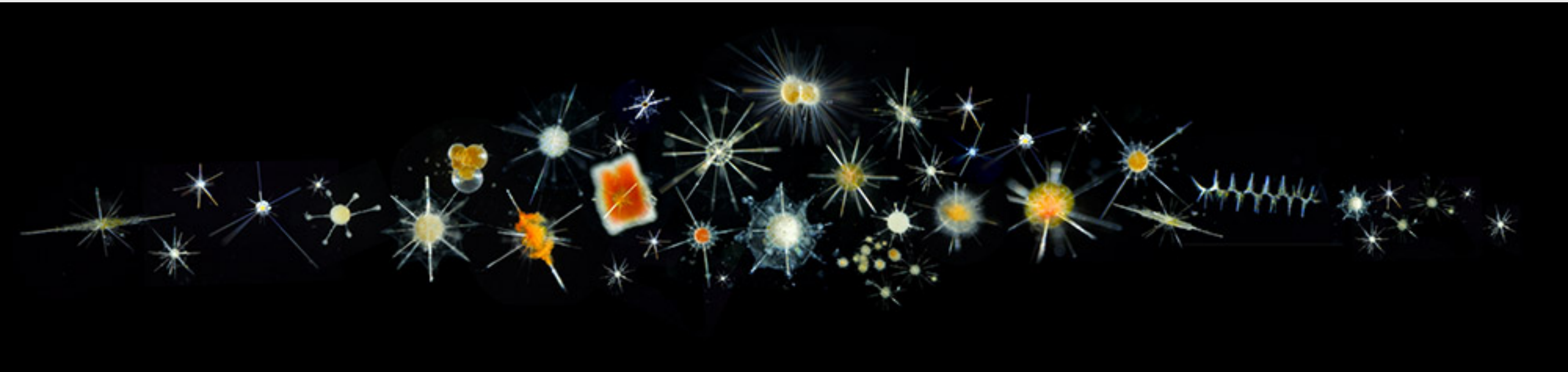
imaginecology2, 2022-09-29
Jean-Olivier Irisson (with input from many colleagues!)

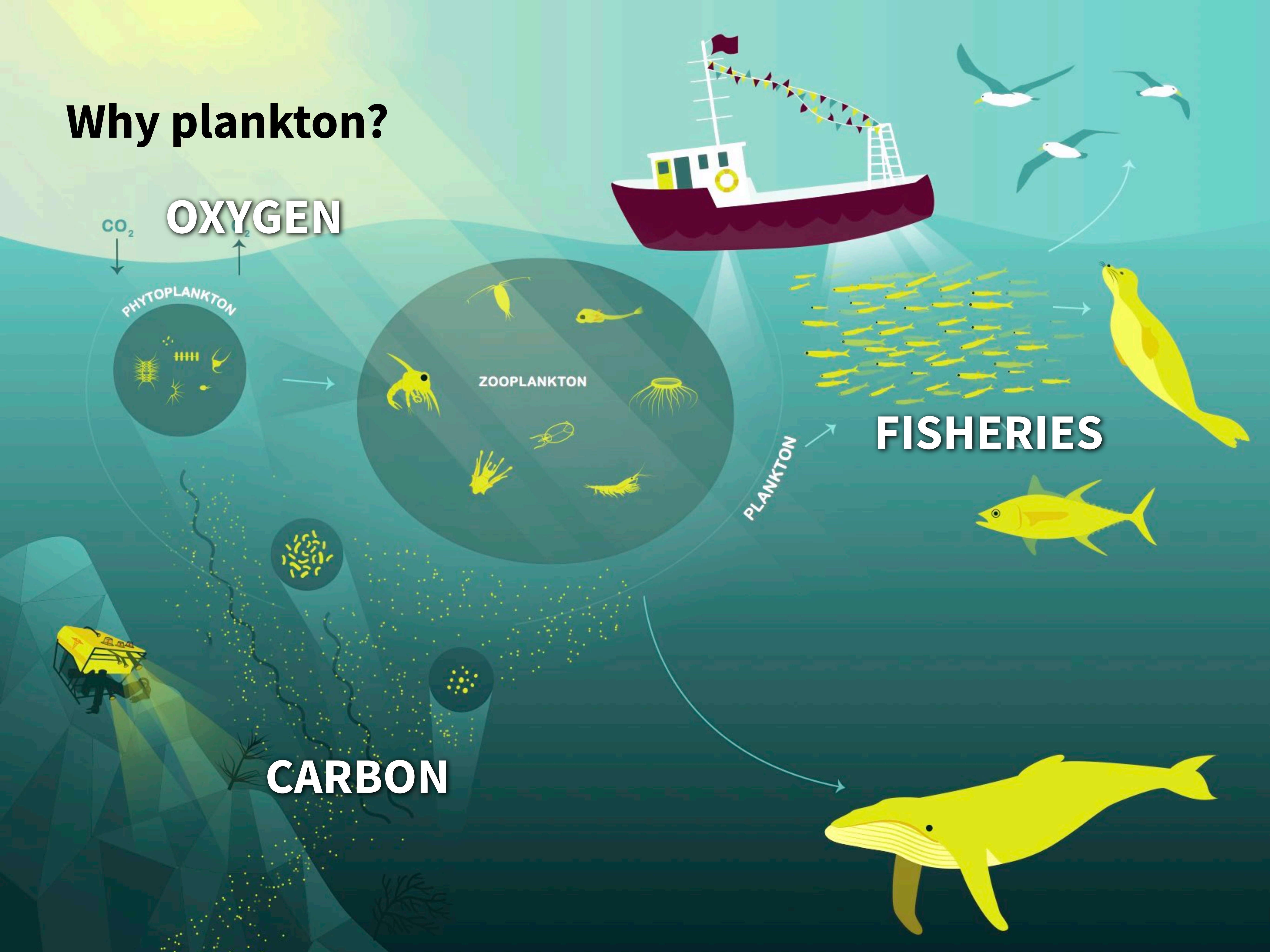
SCIENCES SORBONNE UNIVERSITÉ

Machine learning for plankton and particles images

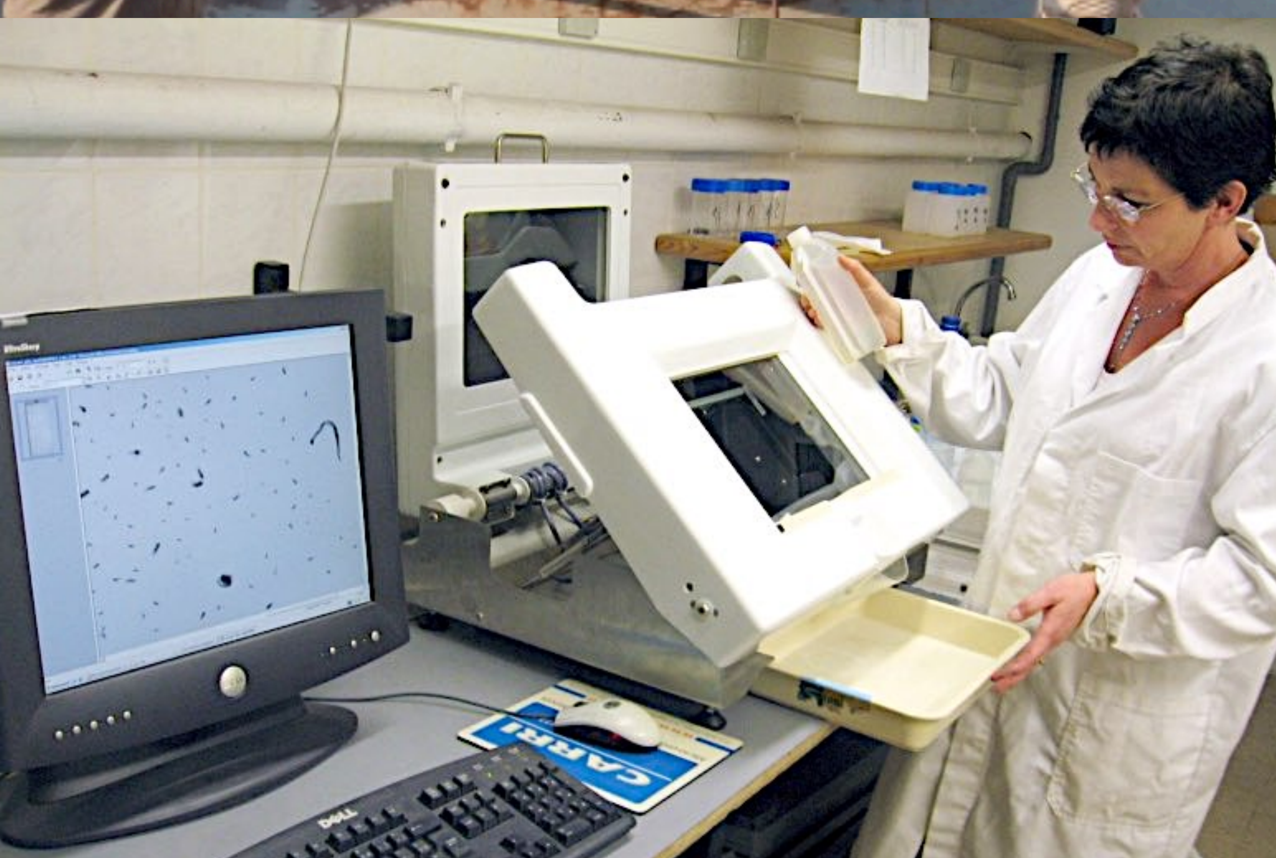
From pictures to data

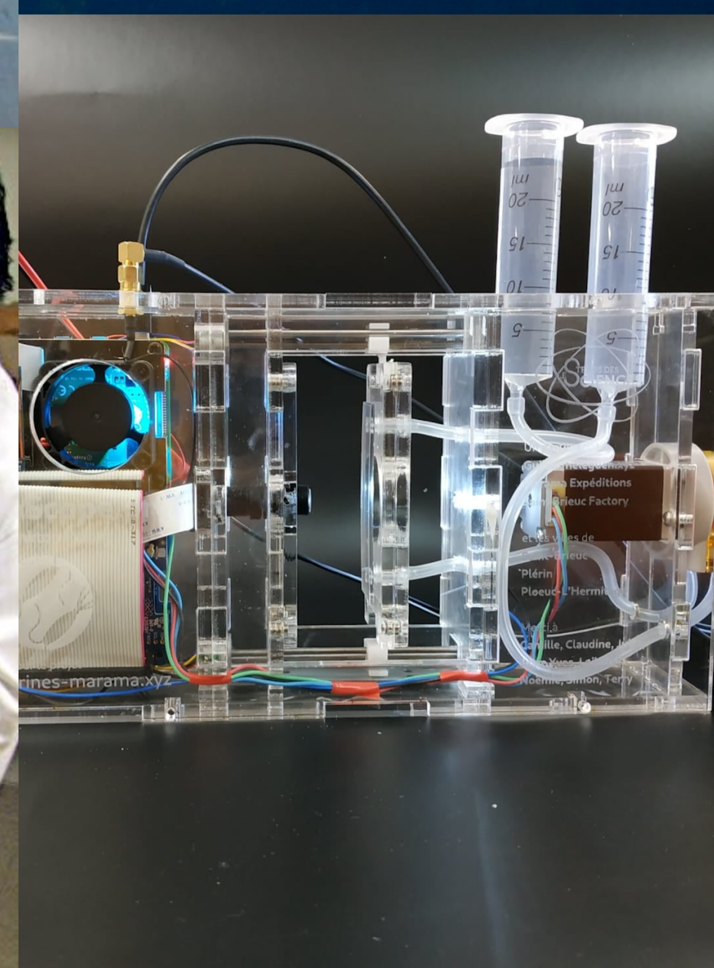


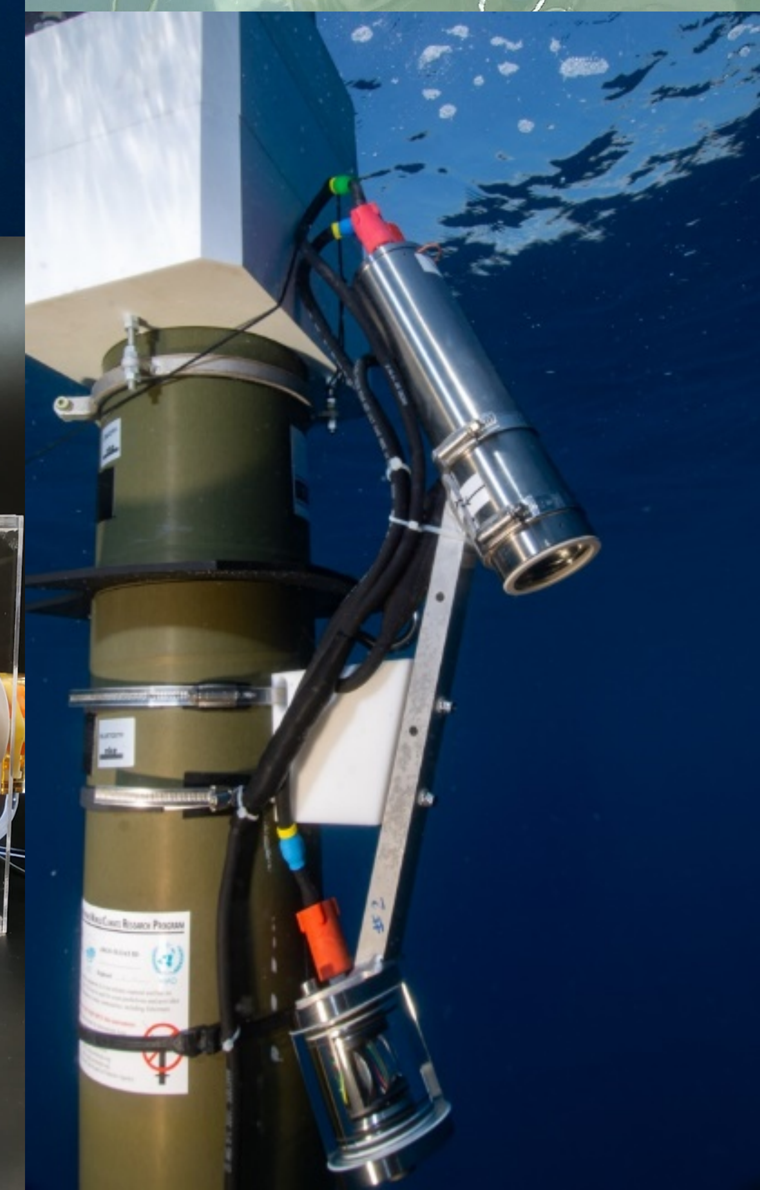
Why plankton?

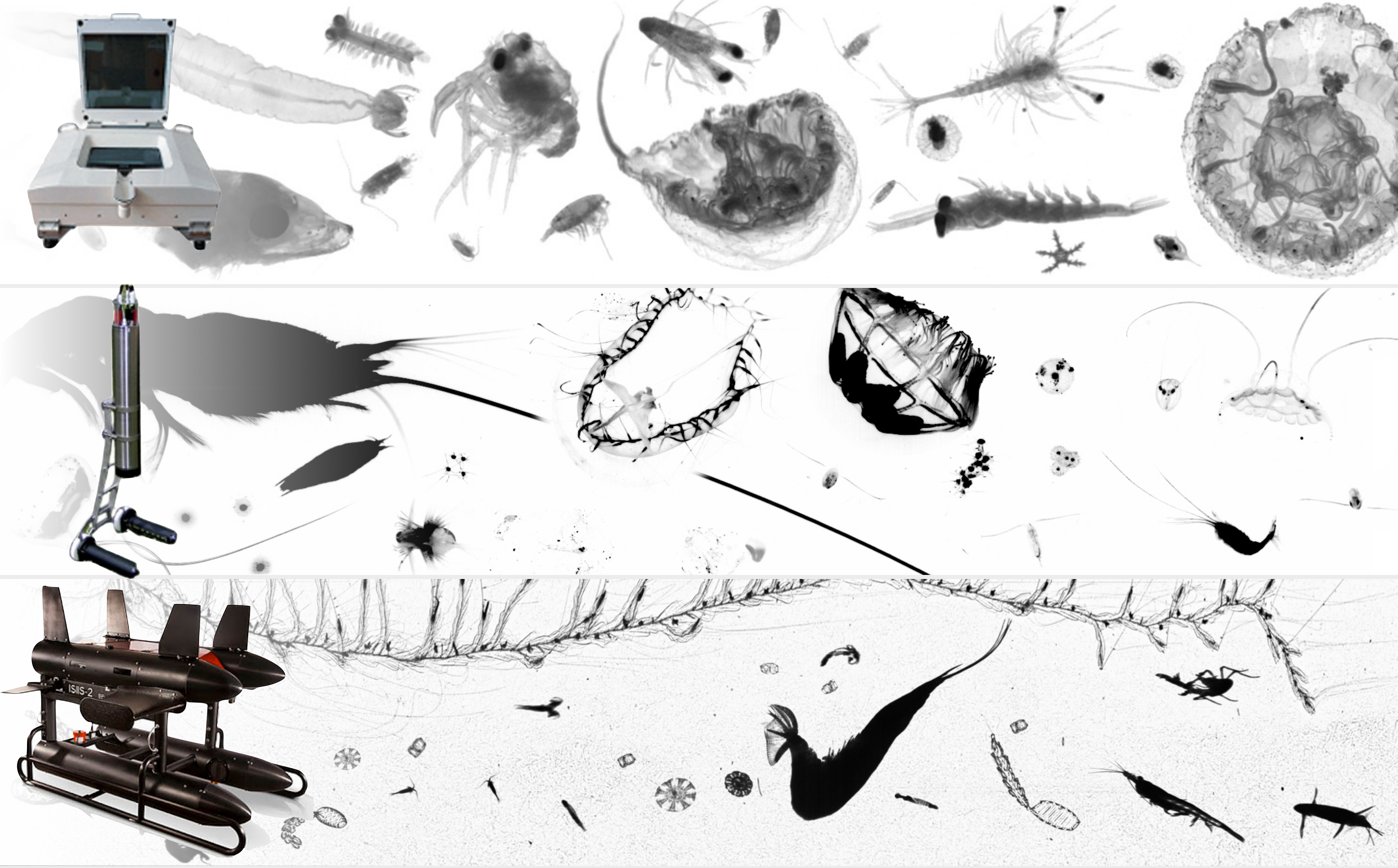


Many instruments





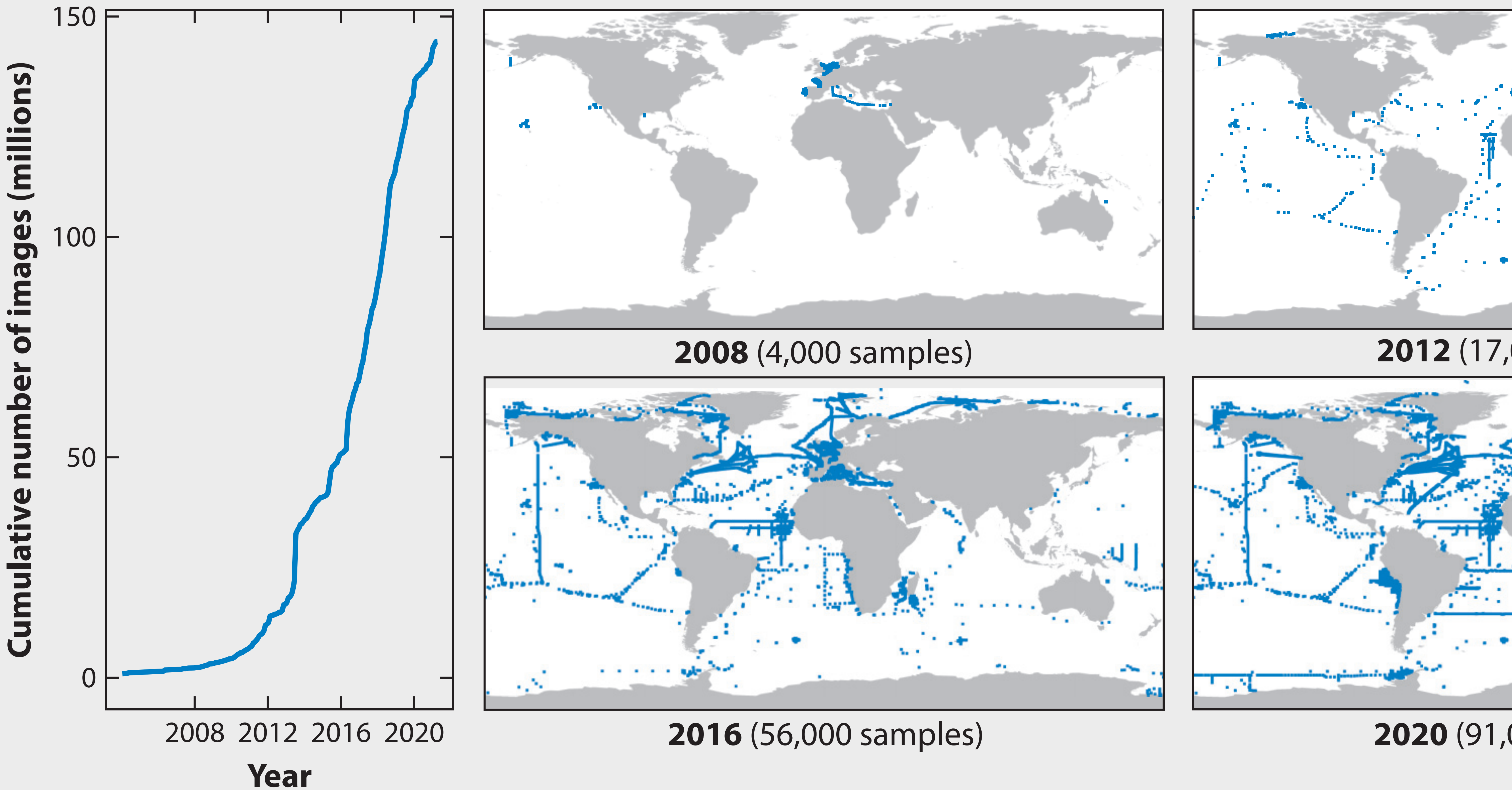




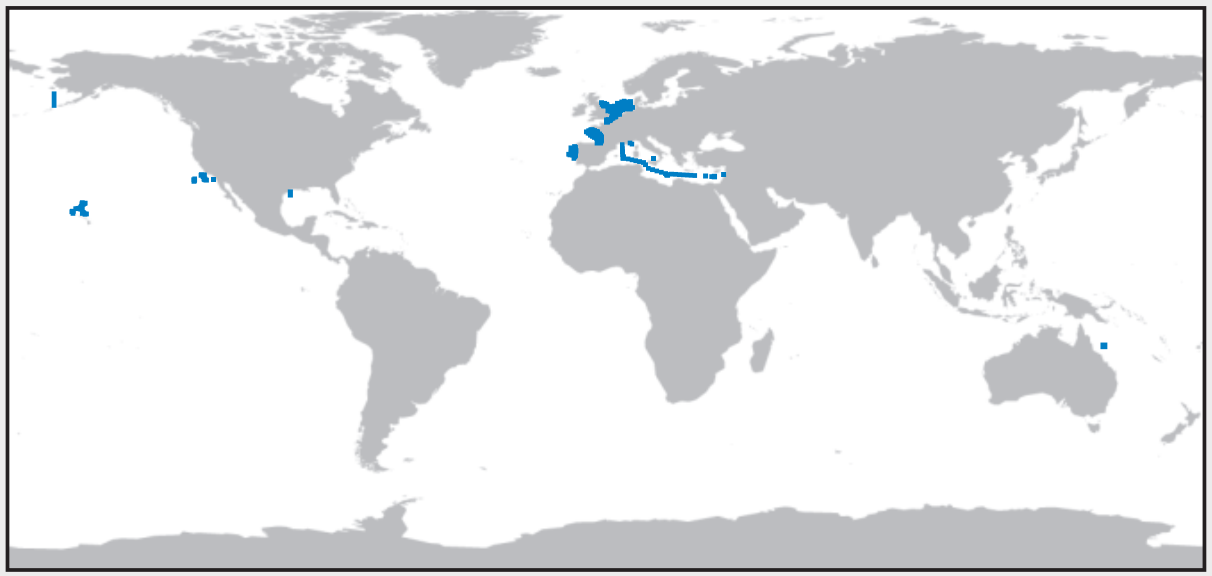
Loads of data

ZooScan = 1 Bpx/y, 1.5M objects/y
UVP = 8.6Bpx/y, ~10M objects/y
ISIIS = 25Tpx/y, 100M objects/y

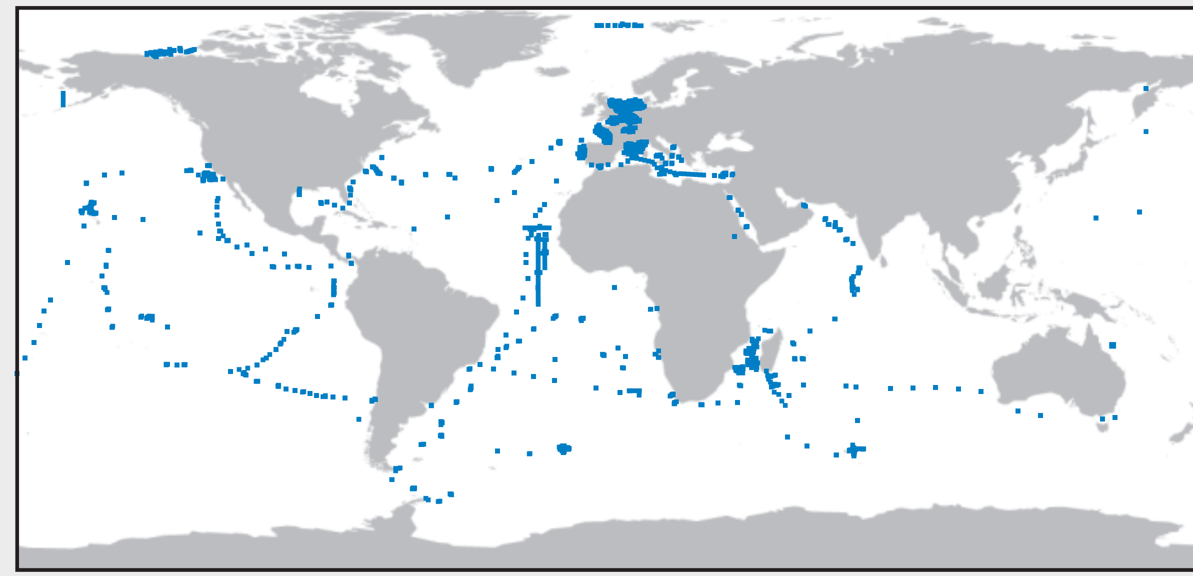
Steep growth in data acquisition



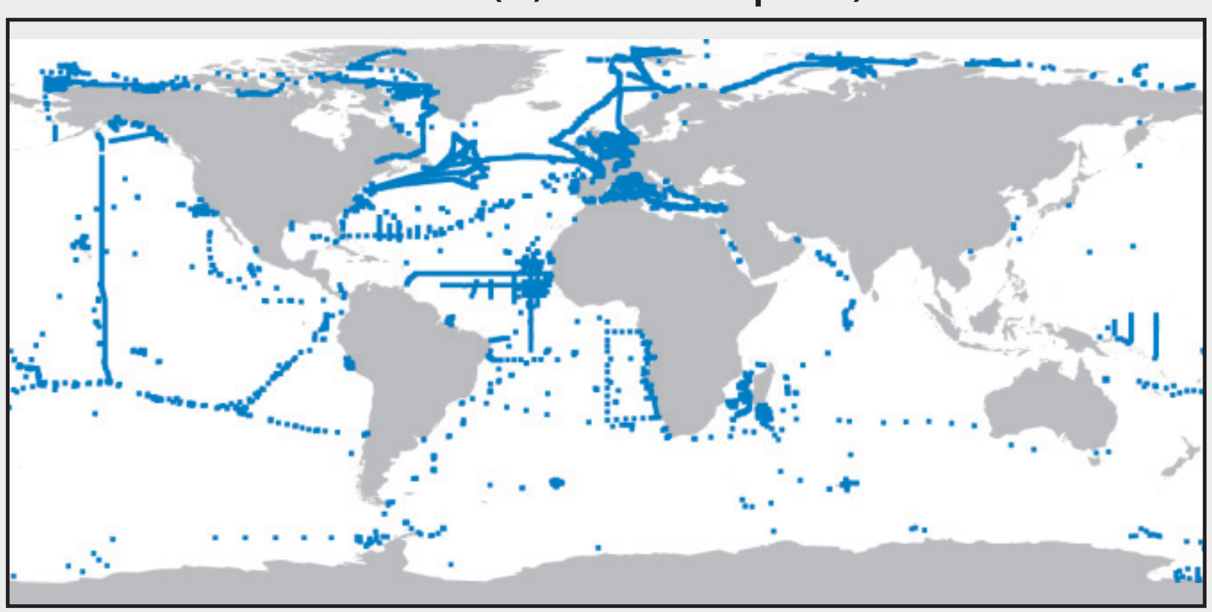
Steep growth in data acquisition



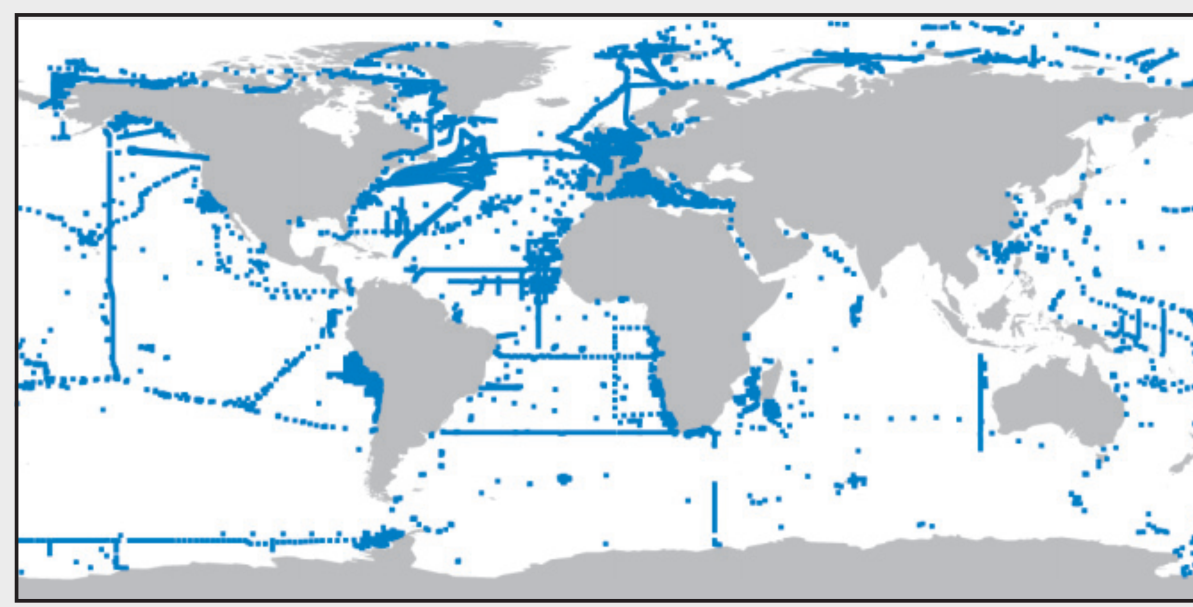
2008 (4,000 samples)



2012 (17,000 samples)

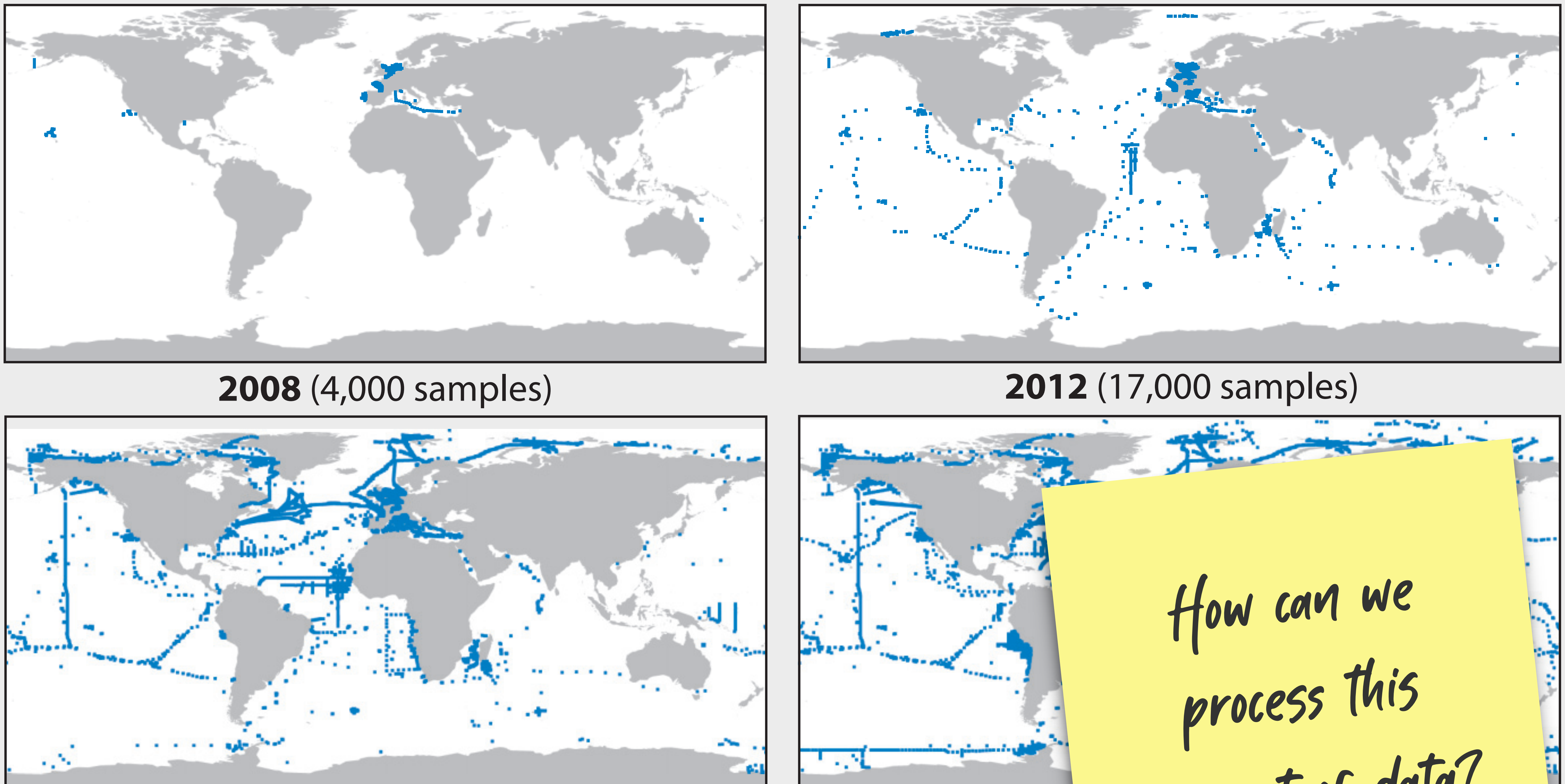


2016 (56,000 samples)



2020 (91,000 samples)

Steep growth in data acquisition



2008 (4,000 samples)

2012 (17,000 samples)

2016 (56,000 samples)

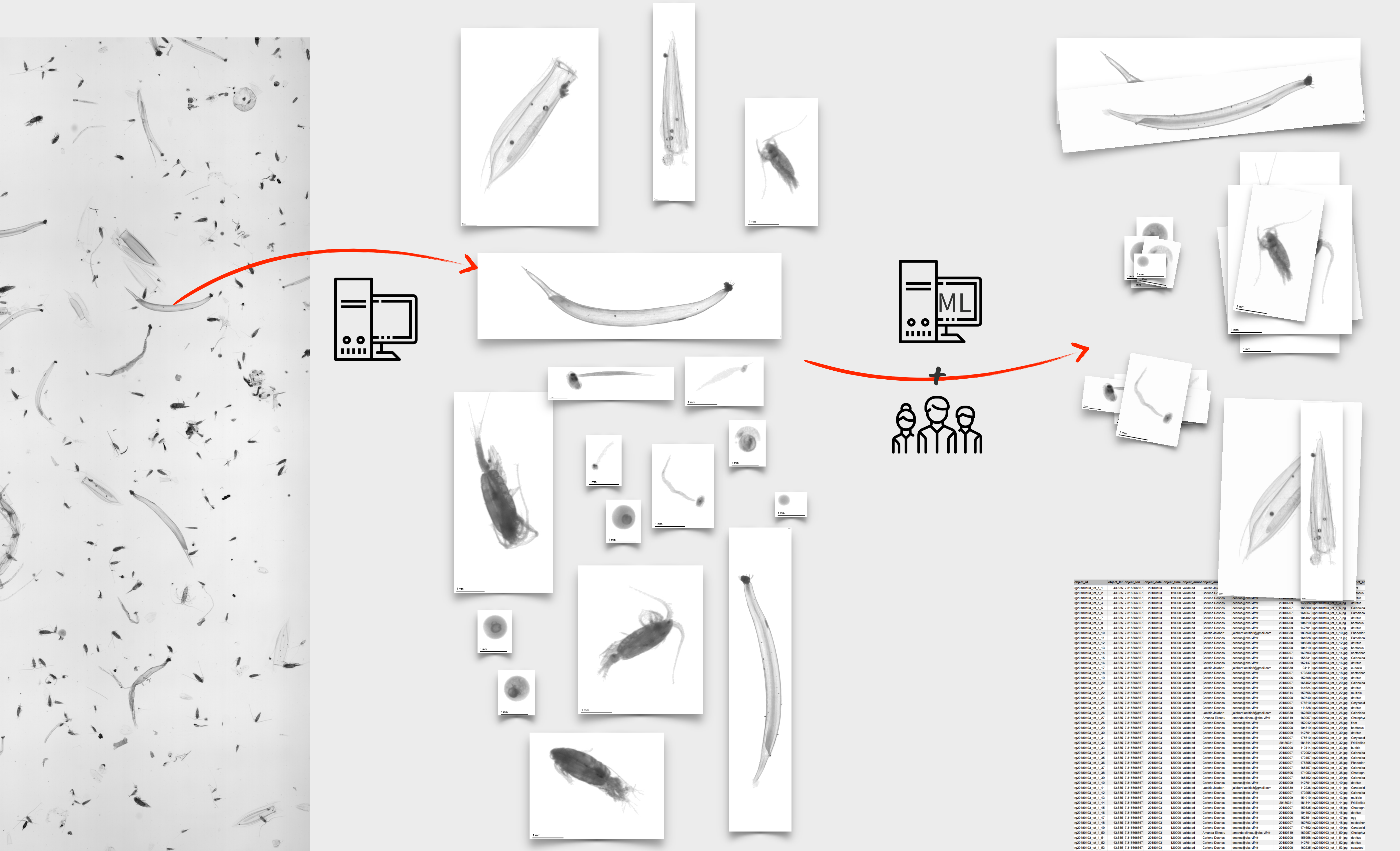
2020

How can we process this amount of data?

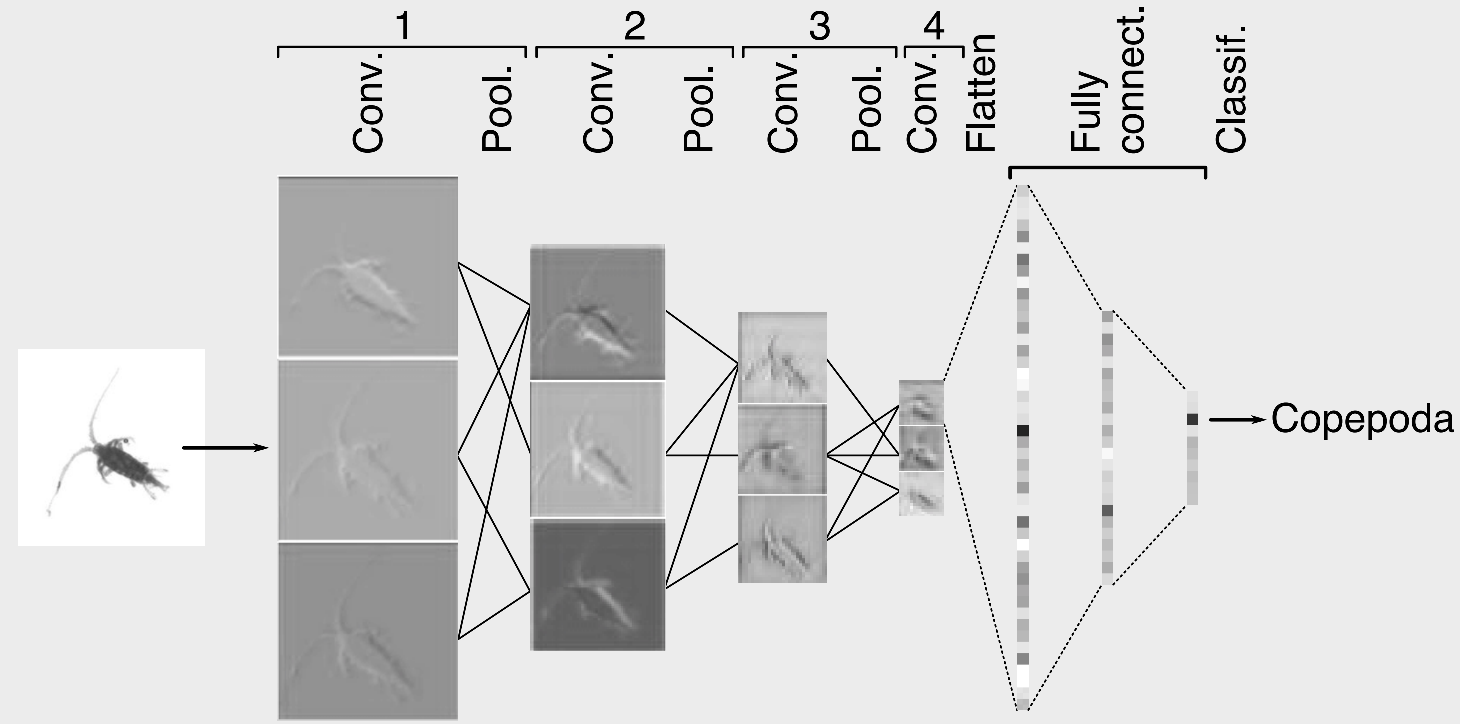
Quantitative imaging and ML-assisted sorting

Quantitative imaging and ML-assisted sorting

Quantitative imaging and ML-assisted sorting



Measure + classify *vs.* **Deep learning**



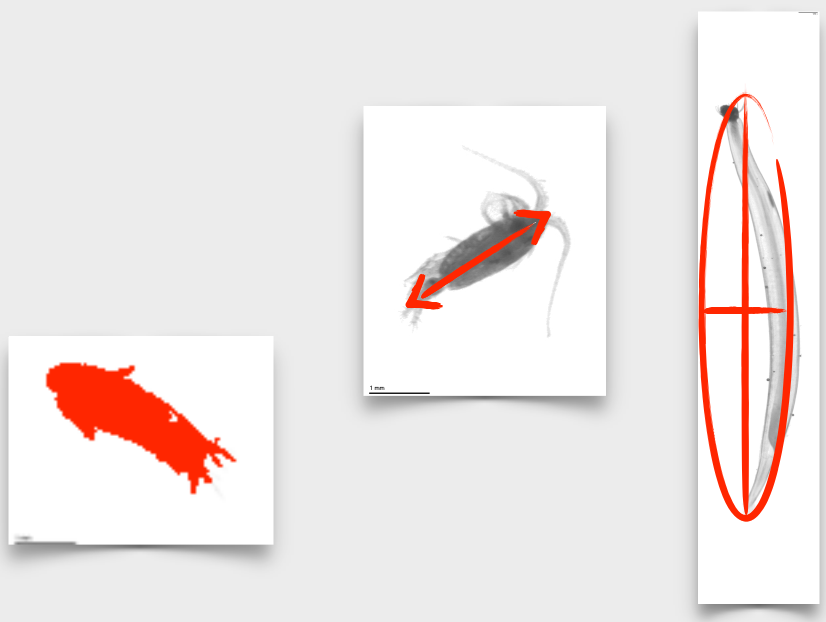
Software to **extract features**
Area (ESD)
Mean/SD of grey
Feret diameter
Major/minor, angle

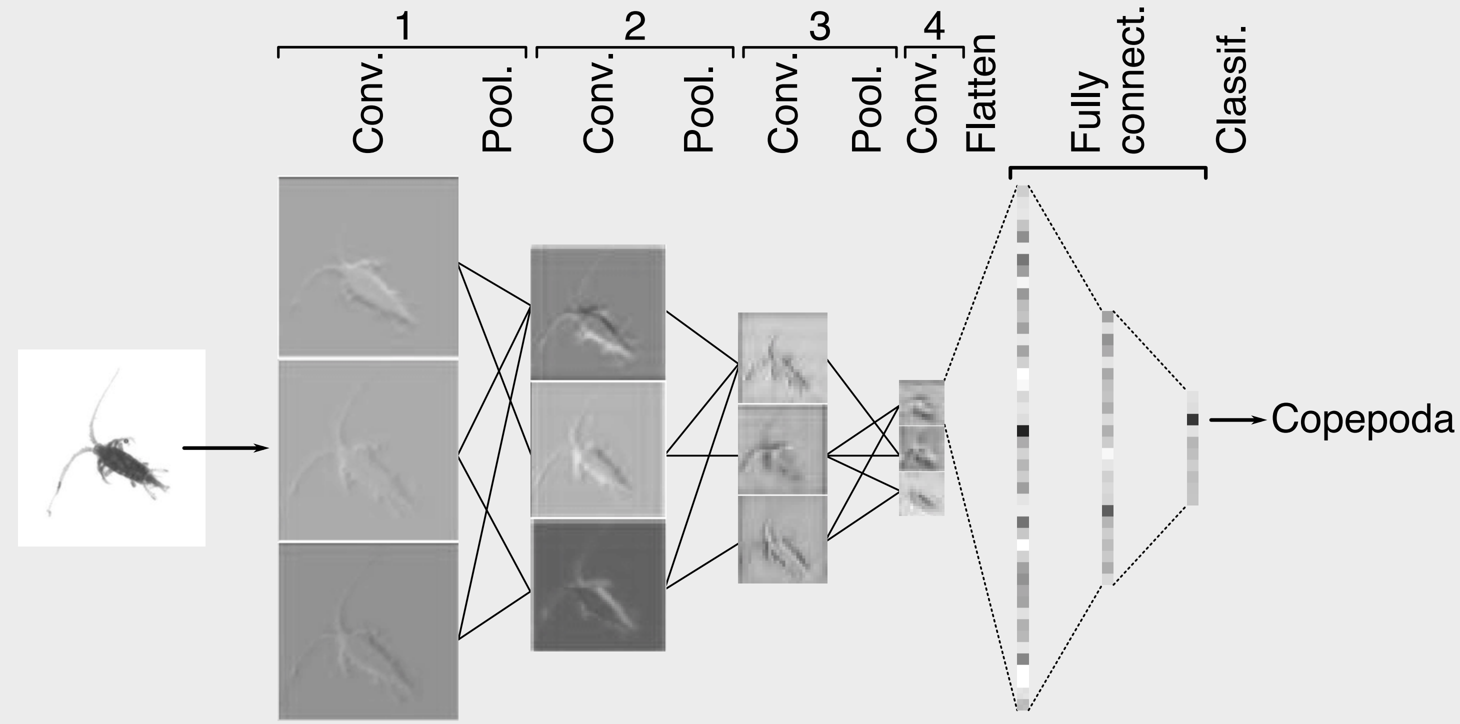
+ a **classifier**

A **feature extractor**
Convolutions
Pooling

+ a **classifier**
Flattening
Fully connected layers

Measure + classify *vs.* Deep learning





Software to **extract features**
Area (ESD)
Mean/SD of grey
Feret diameter
Major/minor, angle

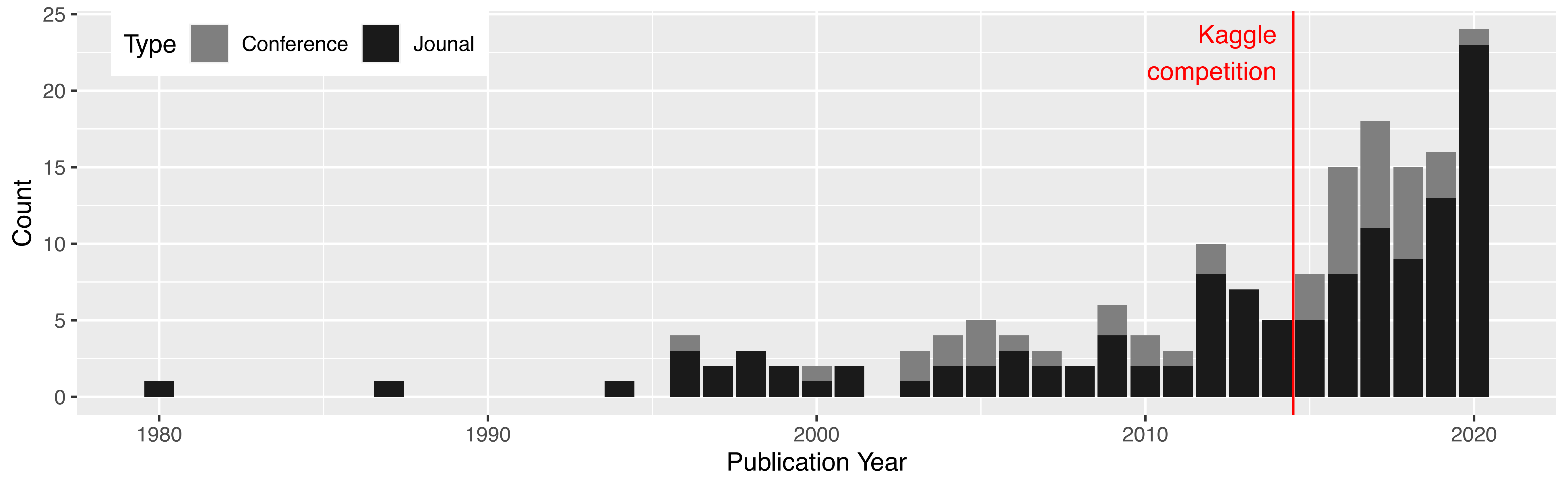
+ a **classifier**

A **feature extractor**
Convolutions
Pooling

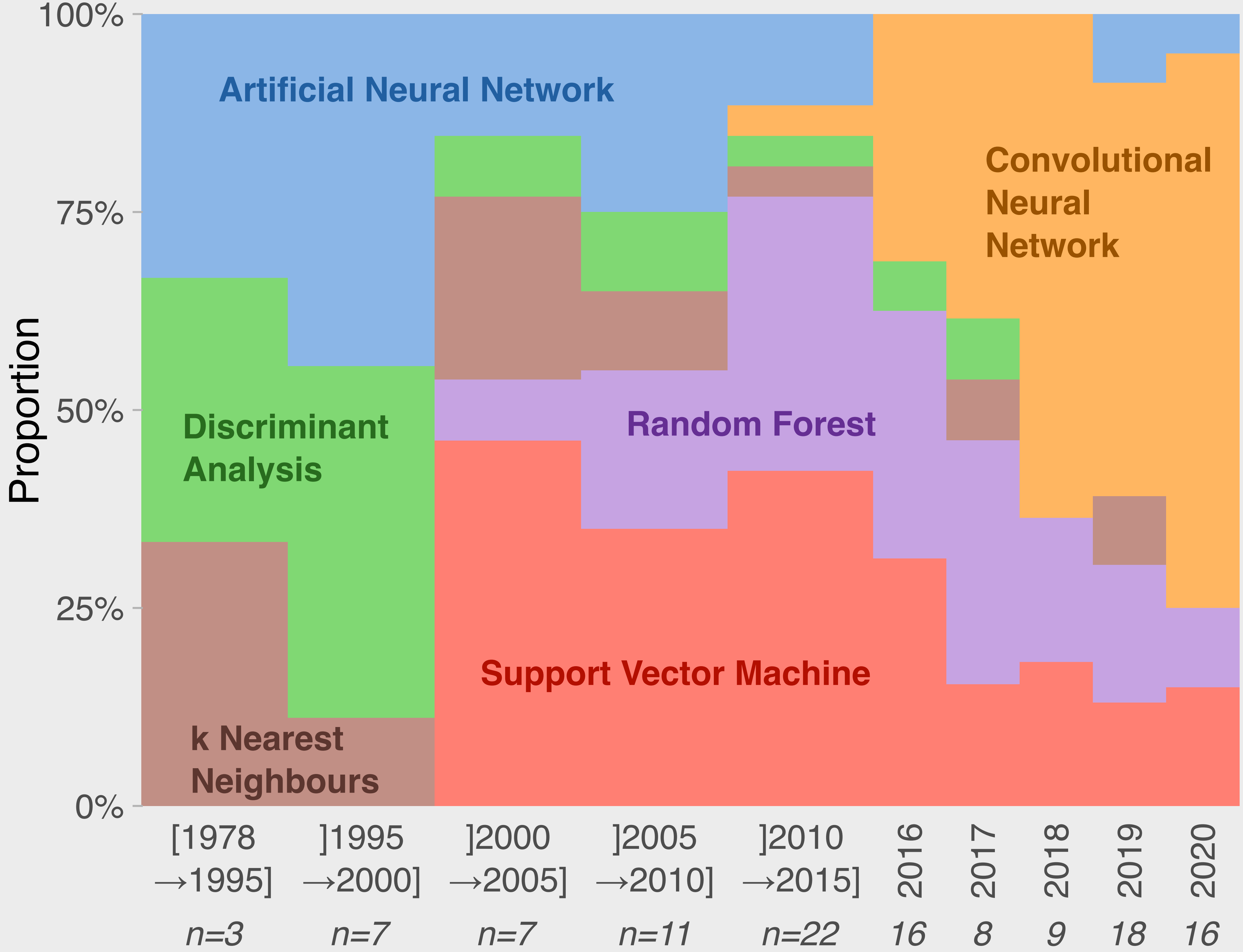
+ a **classifier**
Flattening
Fully connected layers

Plankton image classification is a challenging ML problem

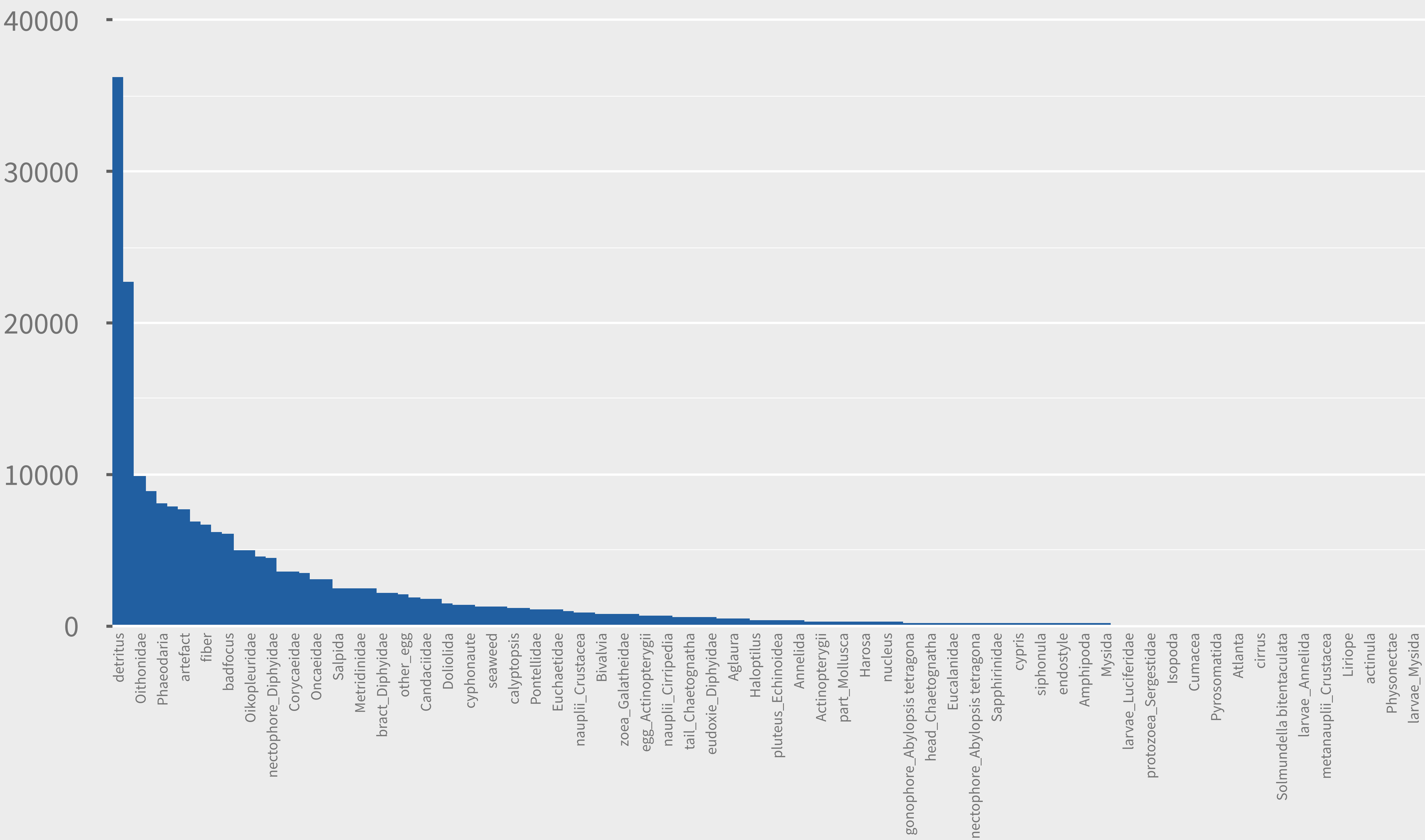
Total: 175 papers!



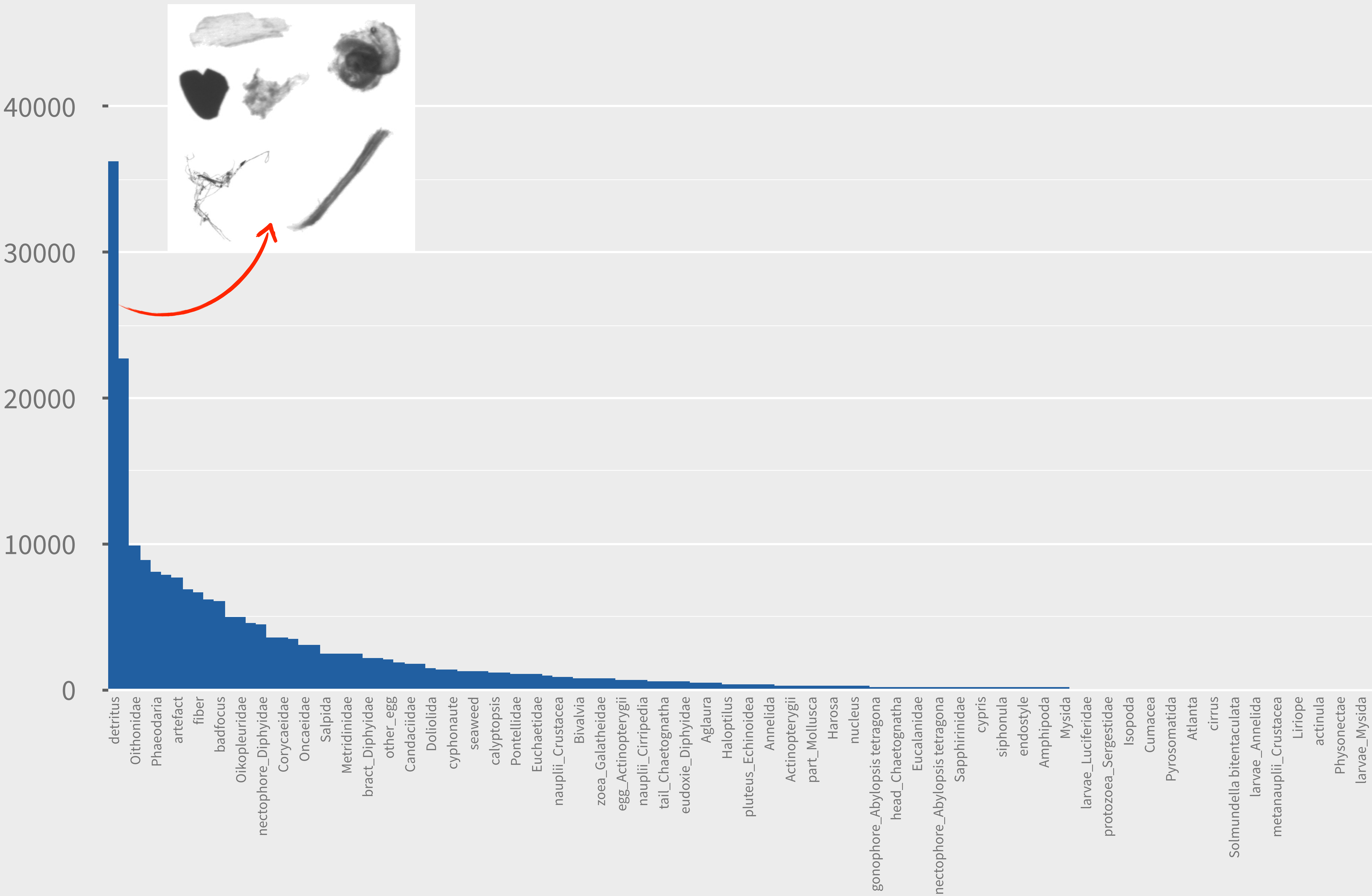
Evolution of machine learning techniques



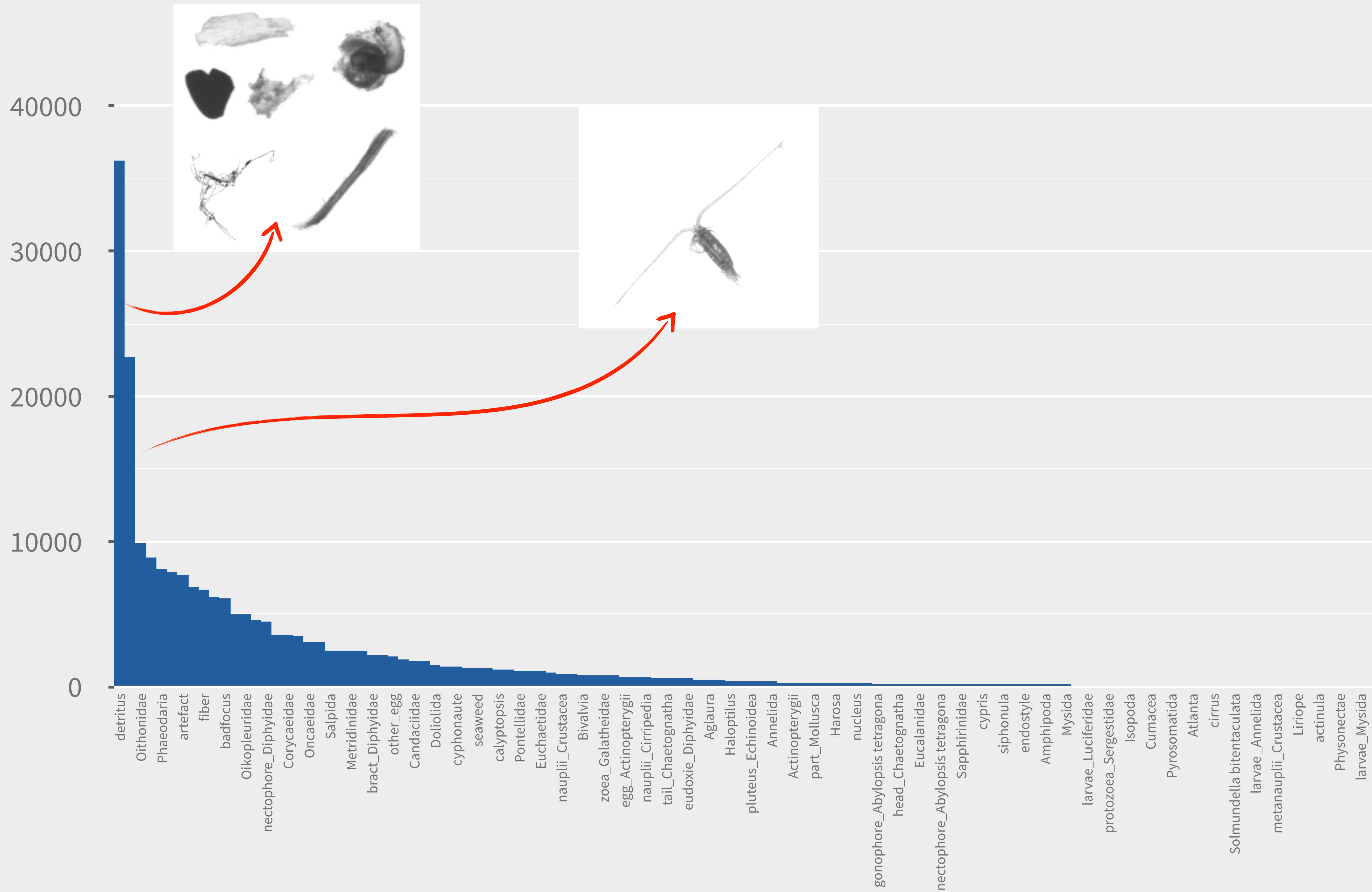
Why is it hard?



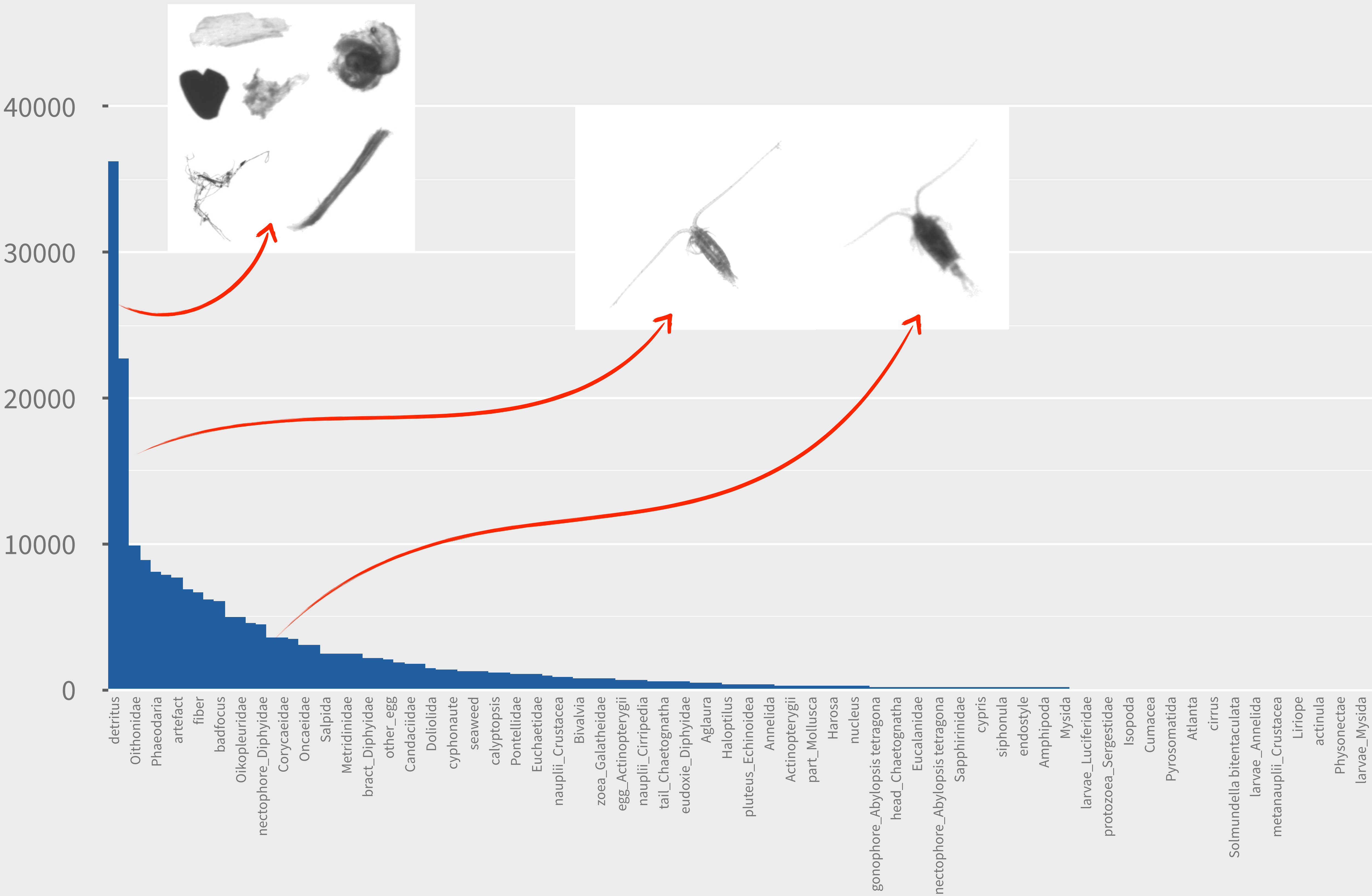
Why is it hard?



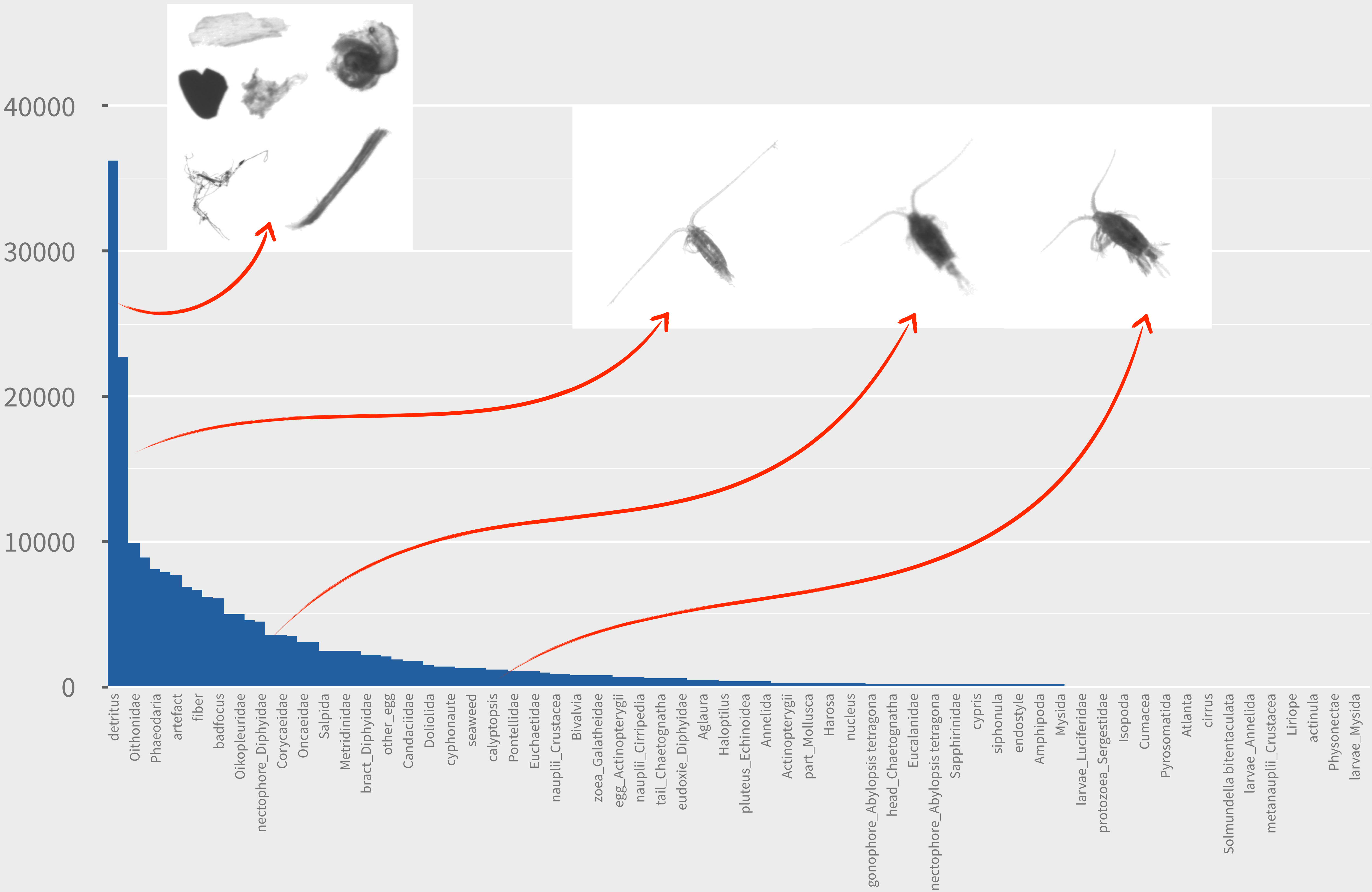
Why is it hard?

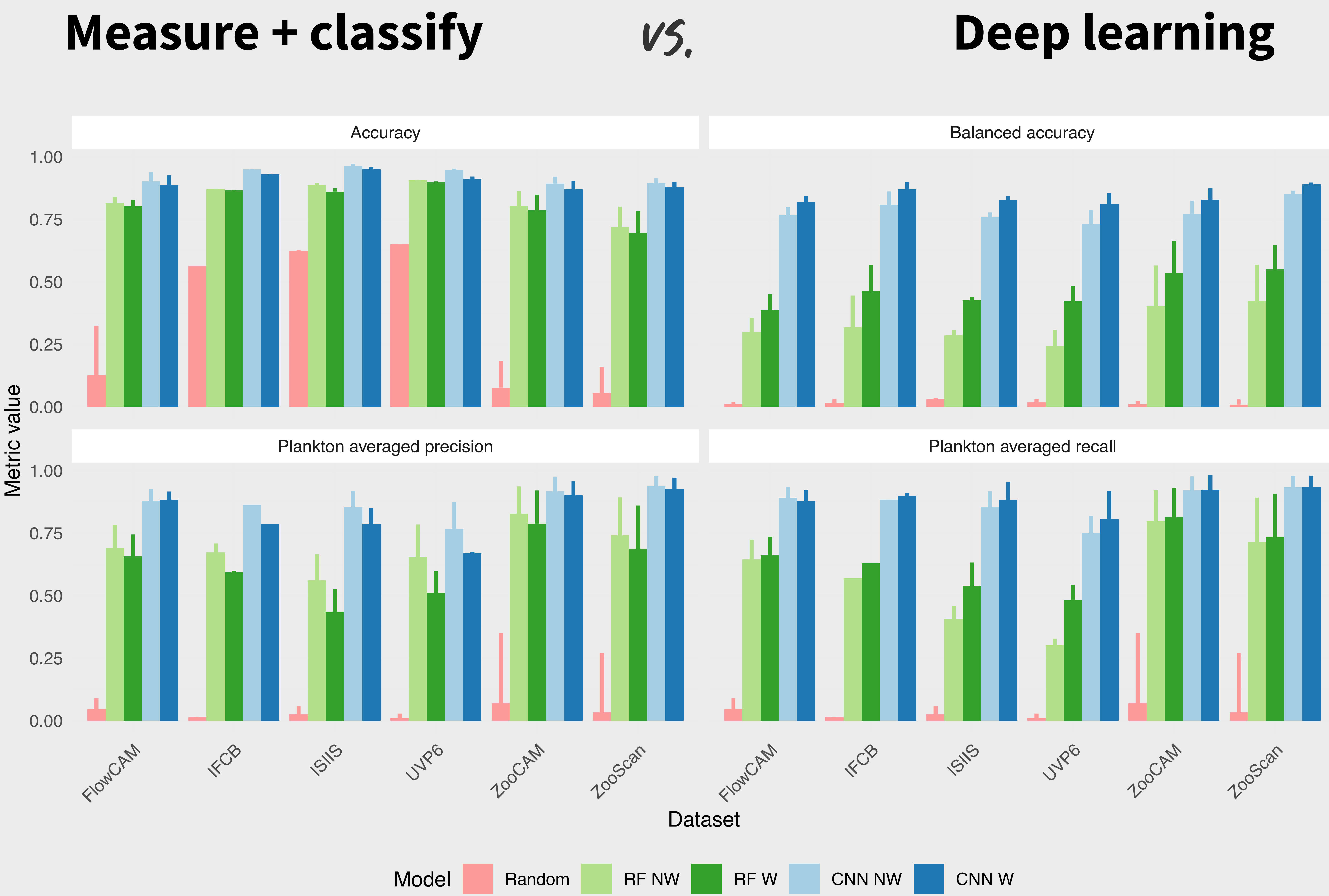


Why is it hard?



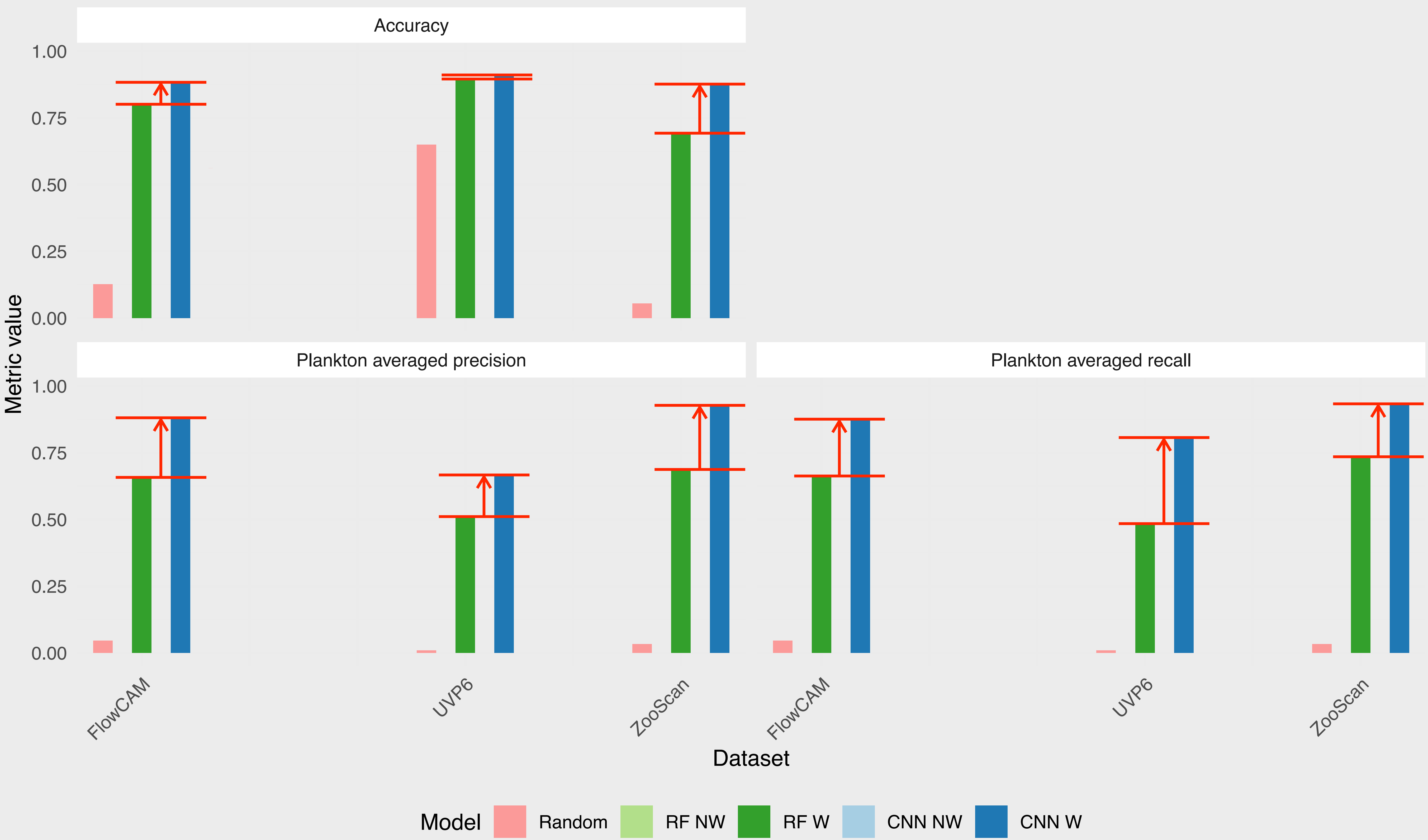
Why is it hard?





Measure + classify
vs.
Deep learning
Accuracy
Balanced accuracy
Plankton averaged precision
Plankton averaged recall
Metric value
1.00
0.75
0.50
0.25
0.00
FlowCAM
IFCB
ISIIS
UVP6
ZooCAM
ZooScan
Dataset
Model
Random
RF NW
RF W
CNN NW
CNN W

Measure + classify *vs.* Deep learning



How deep is enough?

Model	Size	Accuracy	Avg. precision	Avg. recall
MobileNet v4 + 600	5.4M	89.4	91.2	92.0

How deep is enough?

Model	Size	Accuracy	Avg. precision	Avg. recall
MobileNet v4 + 600	5.4M	89.4	91.2	92.0
MobileNet v4 + 1792	7.5M	89.2	90.9	91.9

How deep is enough?

Model	Size	Accuracy	Avg. precision	Avg. recall
MobileNet v4 + 600	5.4M	89.4	91.2	92.0
MobileNet v4 + 1792	7.5M	89.2	90.9	91.9
EfficientNet v2 S + 600	25M	89.8	91.2	92.9

How deep is enough?

Model	Size	Accuracy	Avg. precision	Avg. recall
MobileNet v4 + 600	5.4M	89.4	91.2	92.0
MobileNet v4 + 1792	7.5M	89.2	90.9	91.9
EfficientNet v2 S + 600	25M	89.8	91.2	92.9
EfficientNet v2 XL + 600	208M	89.1	90.9	92.3

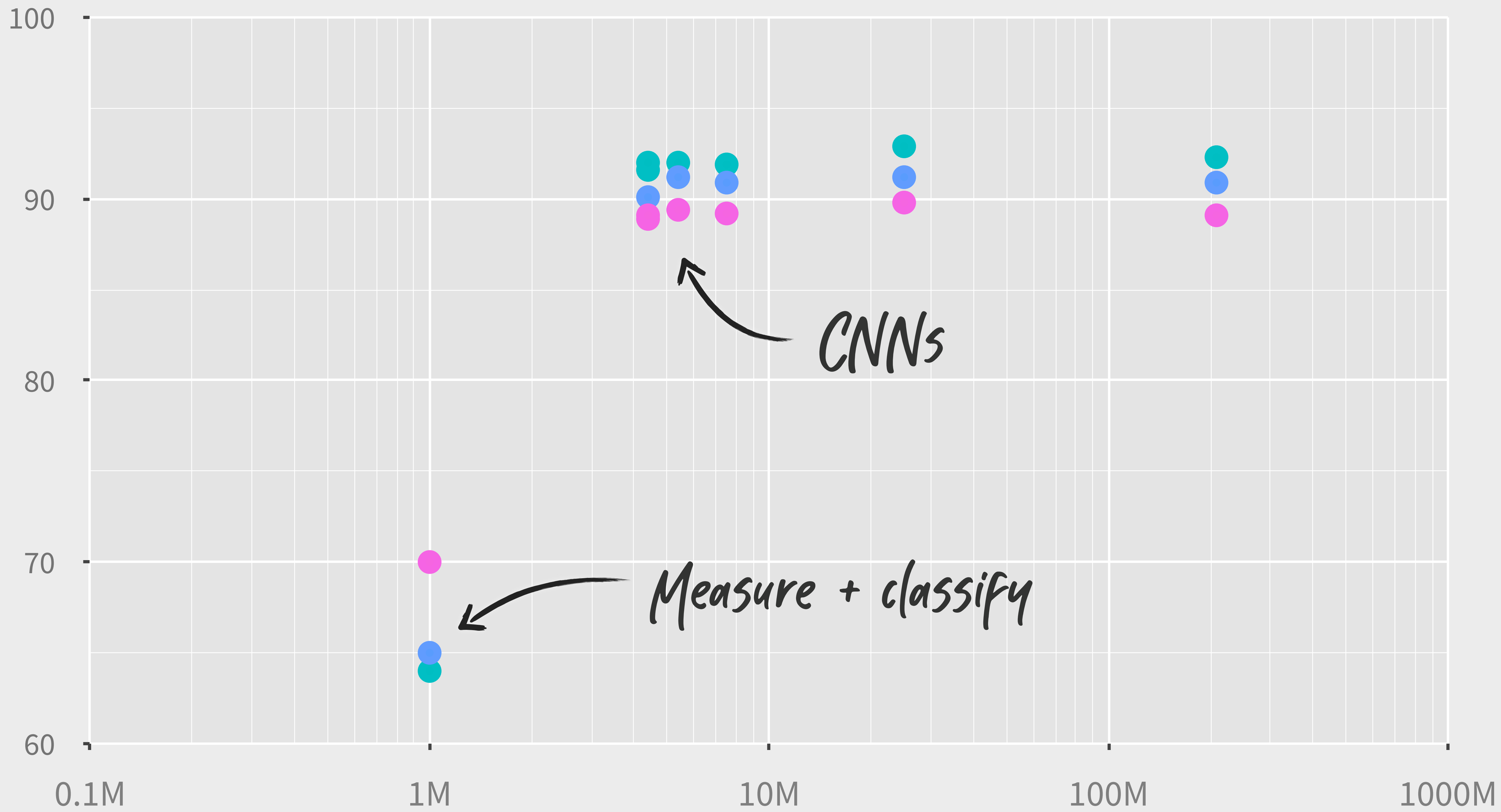
How deep is enough?

Model	Size	Accuracy	Avg. precision	Avg. recall
MobileNet v4 + 600	5.4M	89.4	91.2	92.0
MobileNet v4 + 1792	7.5M	89.2	90.9	91.9
EfficientNet v2 S + 600	25M	89.8	91.2	92.9
EfficientNet v2 XL + 600	208M	89.1	90.9	92.3
MobileNet v4 + 50	4.4M	88.9	90.1	901.6

How deep is enough?

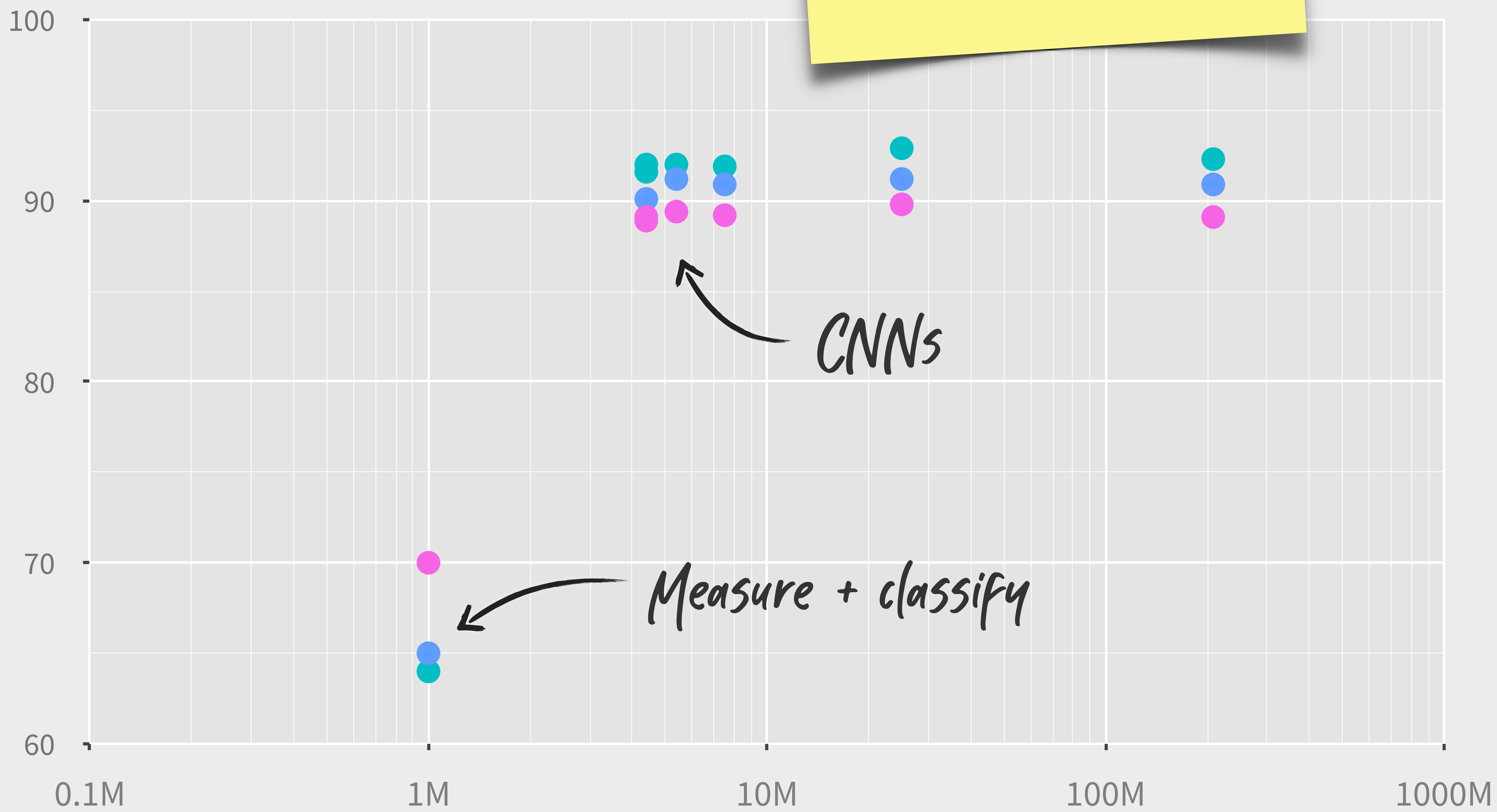
Model	Size	Accuracy	Avg. precision	Avg. recall
MobileNet v4 + 600	5.4M	89.4	91.2	92.0
MobileNet v4 + 1792	7.5M	89.2	90.9	91.9
EfficientNet v2 S + 600	25M	89.8	91.2	92.9
EfficientNet v2 XL + 600	208M	89.1	90.9	92.3
MobileNet v4 + 50	4.4M	88.9	90.1	901.6
MobileNet v4 features + PCA + RF	~4.4M	89.1	90.1	92.0

How deep is enough?



How deep is enough?

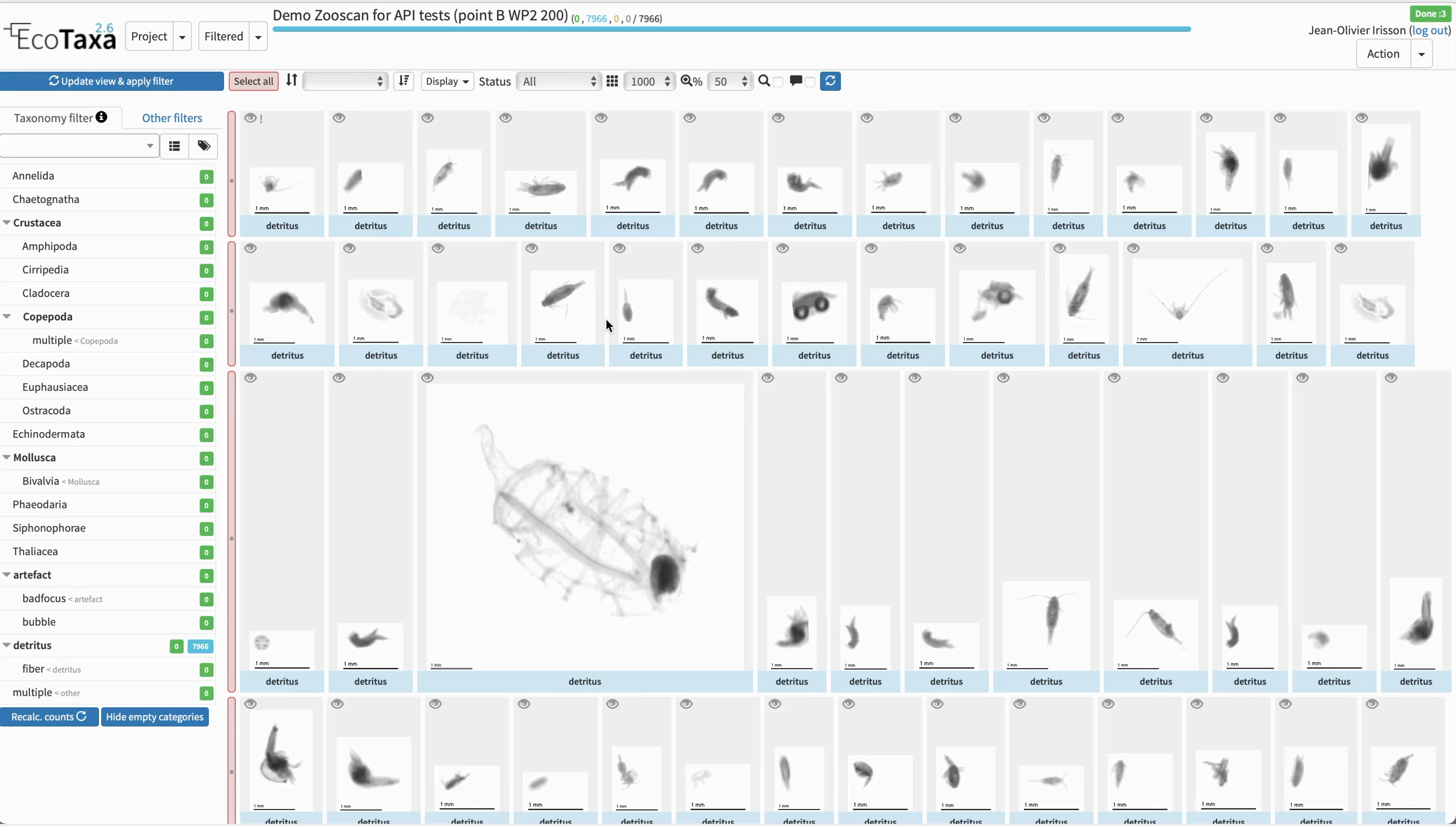
For plankton images:
not very deep



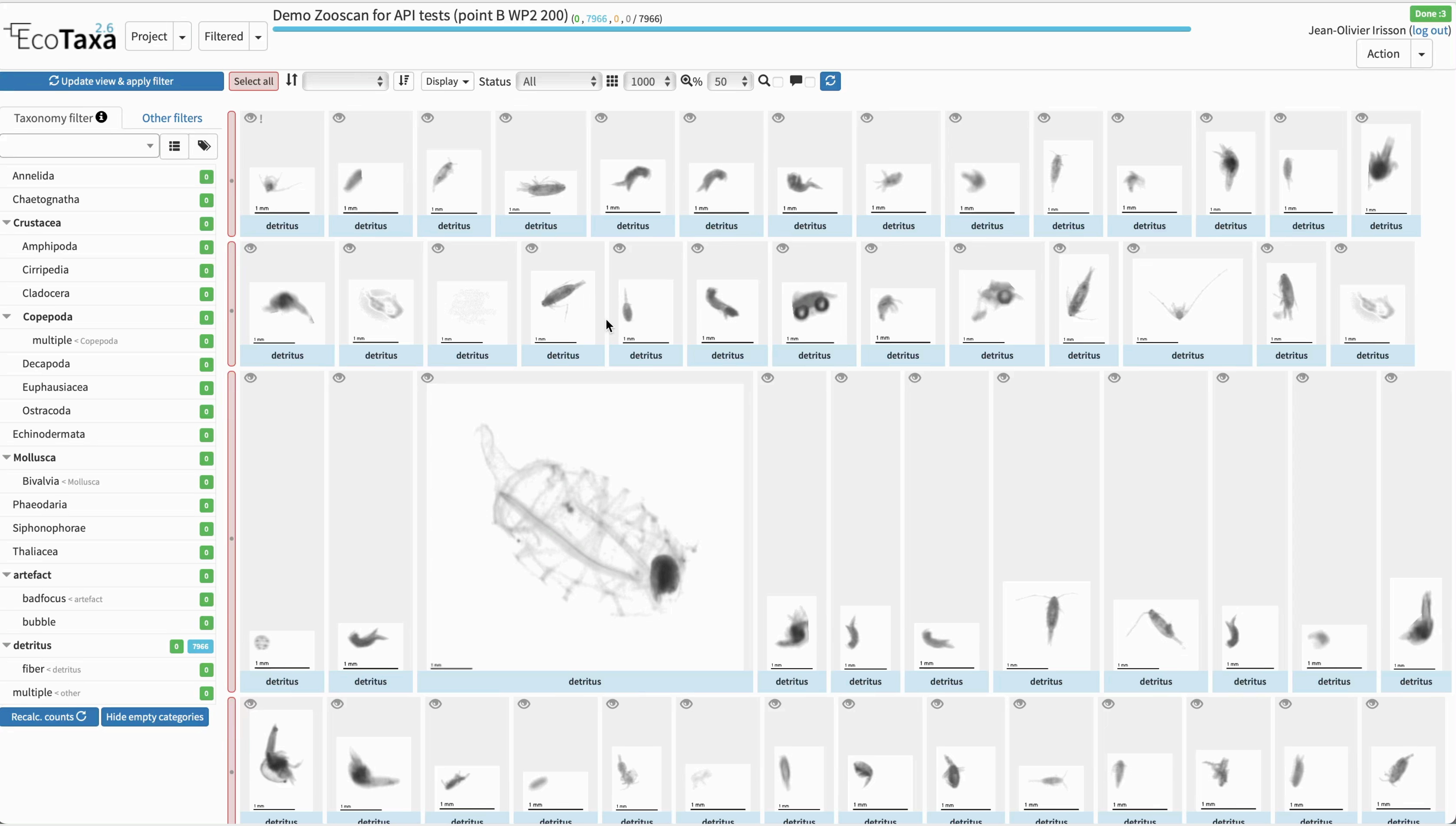
And in real life?

Performance metrics are hard to interpret!

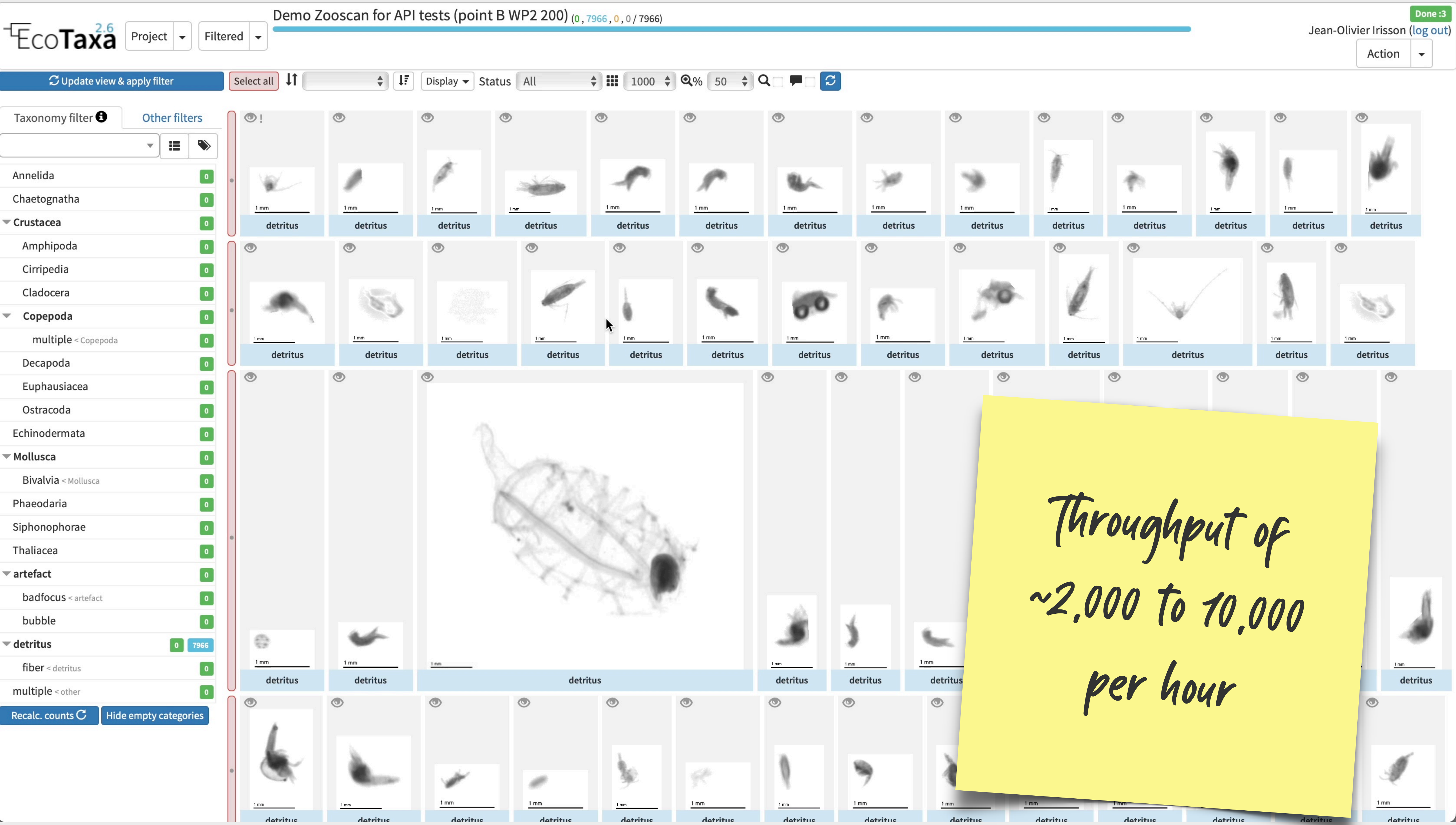
EcoTaxa: ML-assisted image classification



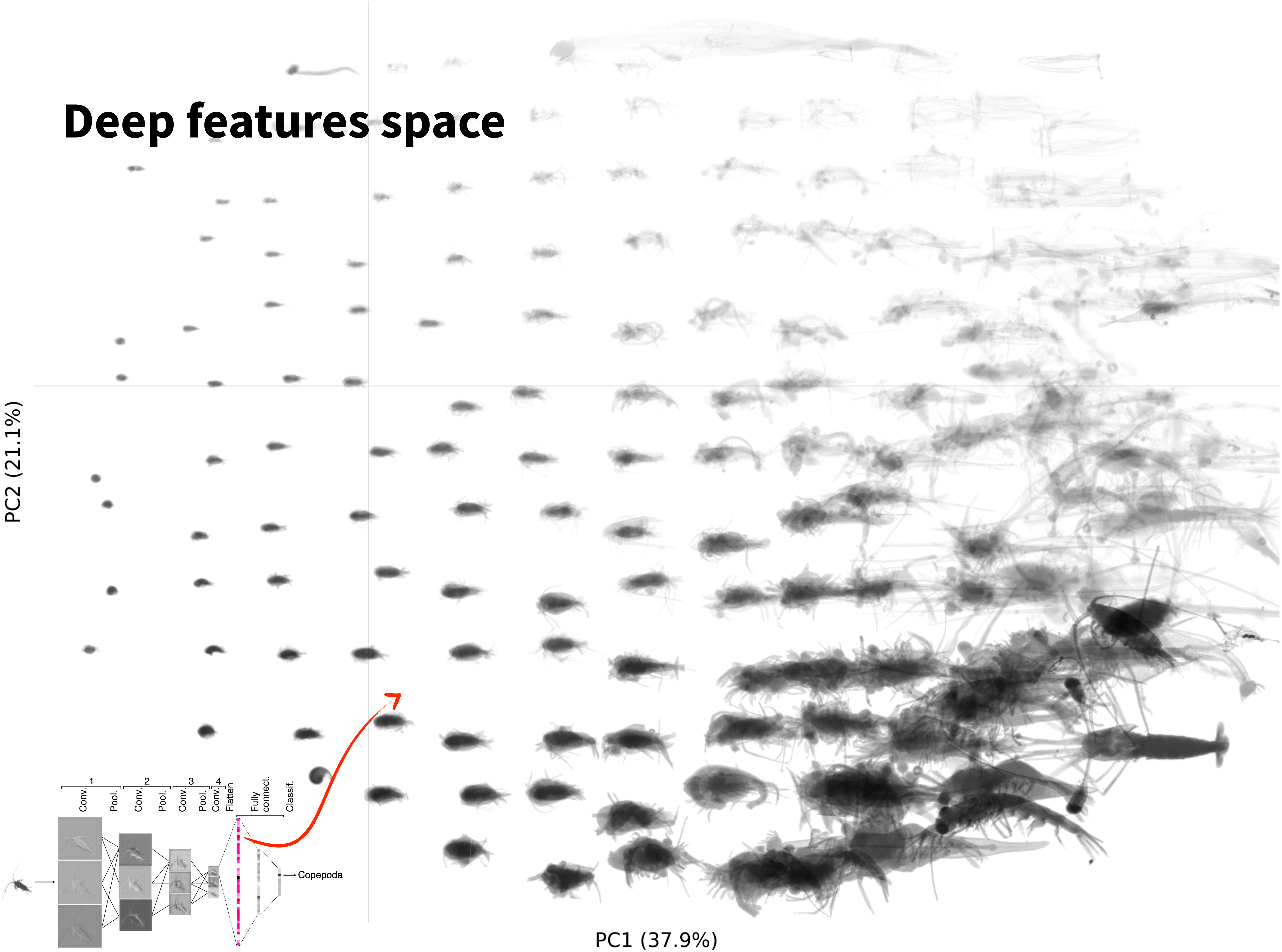
EcoTaxa: ML-assisted image classification



EcoTaxa: ML-assisted image classification



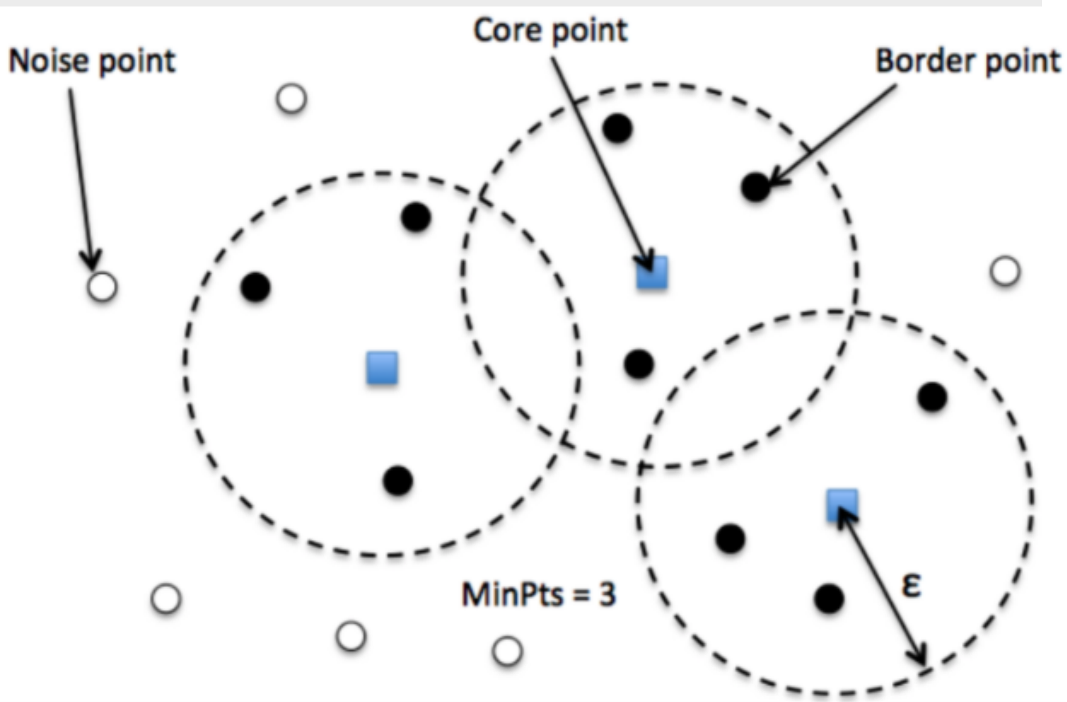
Deep features space

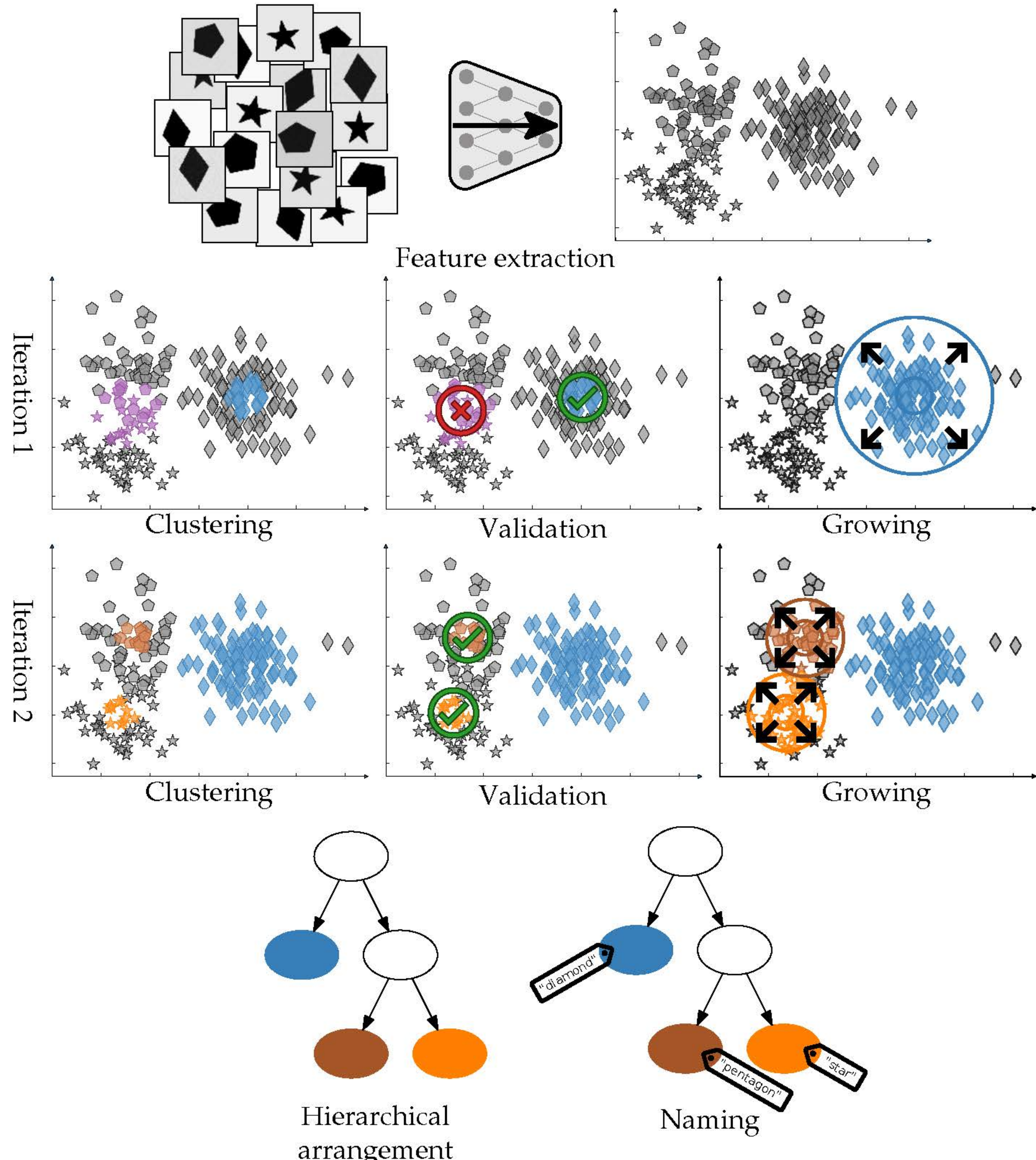


Unsupervised classification

1. Density based clustering to find seeds
2. Growing
3. Labelling

MorphoCluster

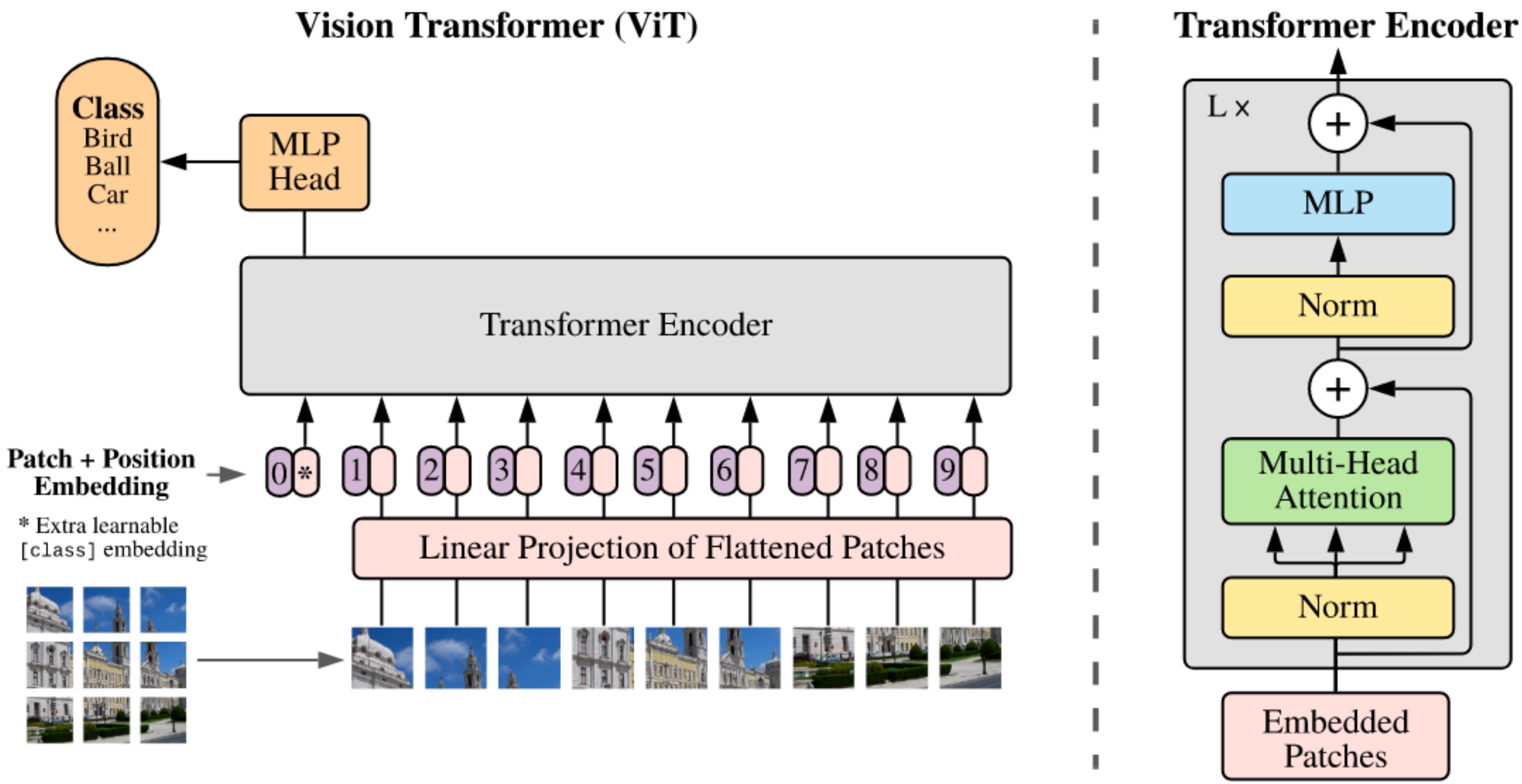




What’s next?

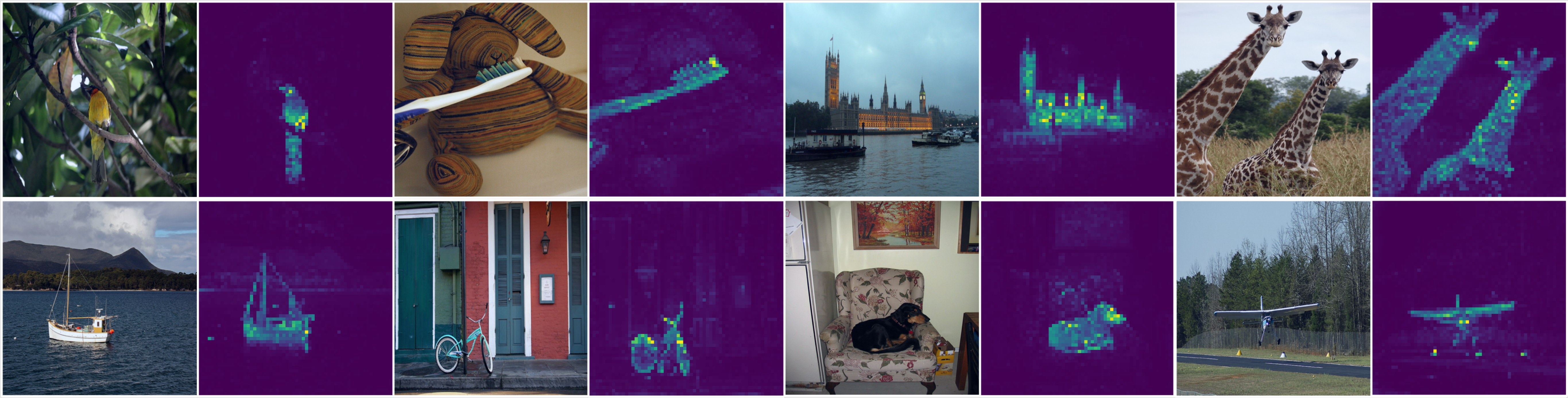
What’s next?

Vision transformers and self-supervision



Vision transformers and self-supervision

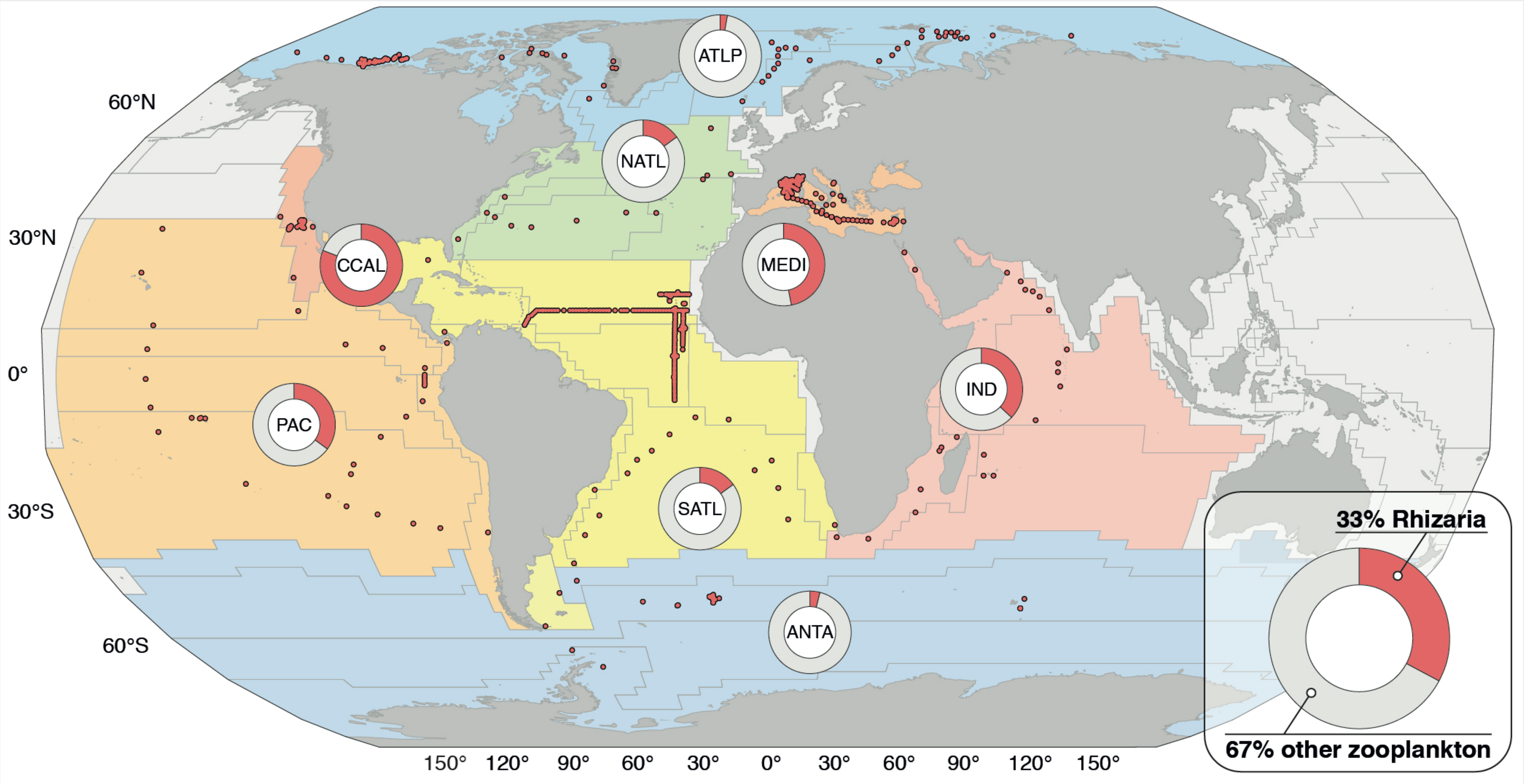
Vision transformers and self-supervision




Merci

What for?

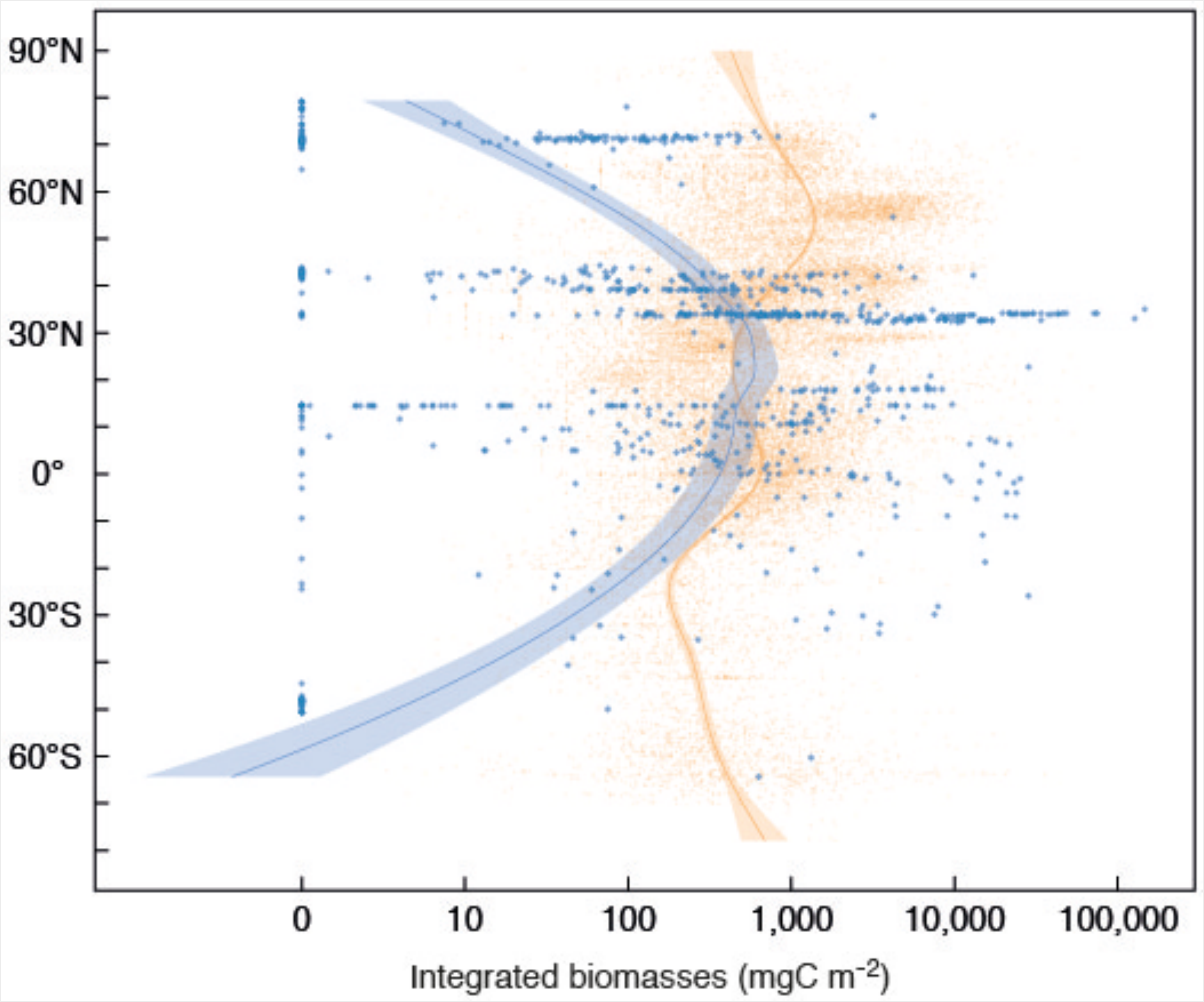
Global biomass of fragile plankton



Biard T, Stemmann L, Picheral M, Mayot N, Vandromme P, Hauss H, Gorsky G, Guidi L, Kiko R, Not F (2016)
In situ imaging reveals the biomass of giant protists in the global ocean. Nature 532:504.

What for?

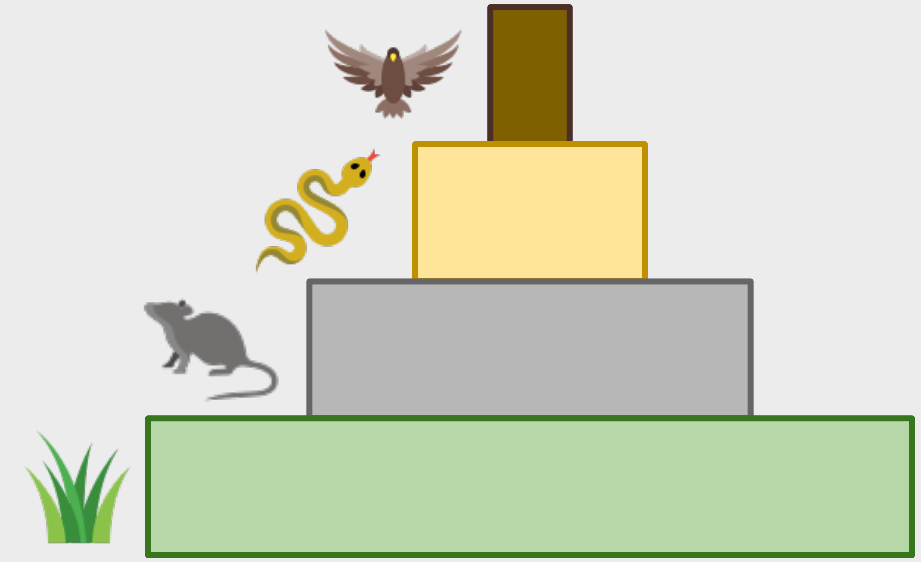
Global biomass of fragile plankton

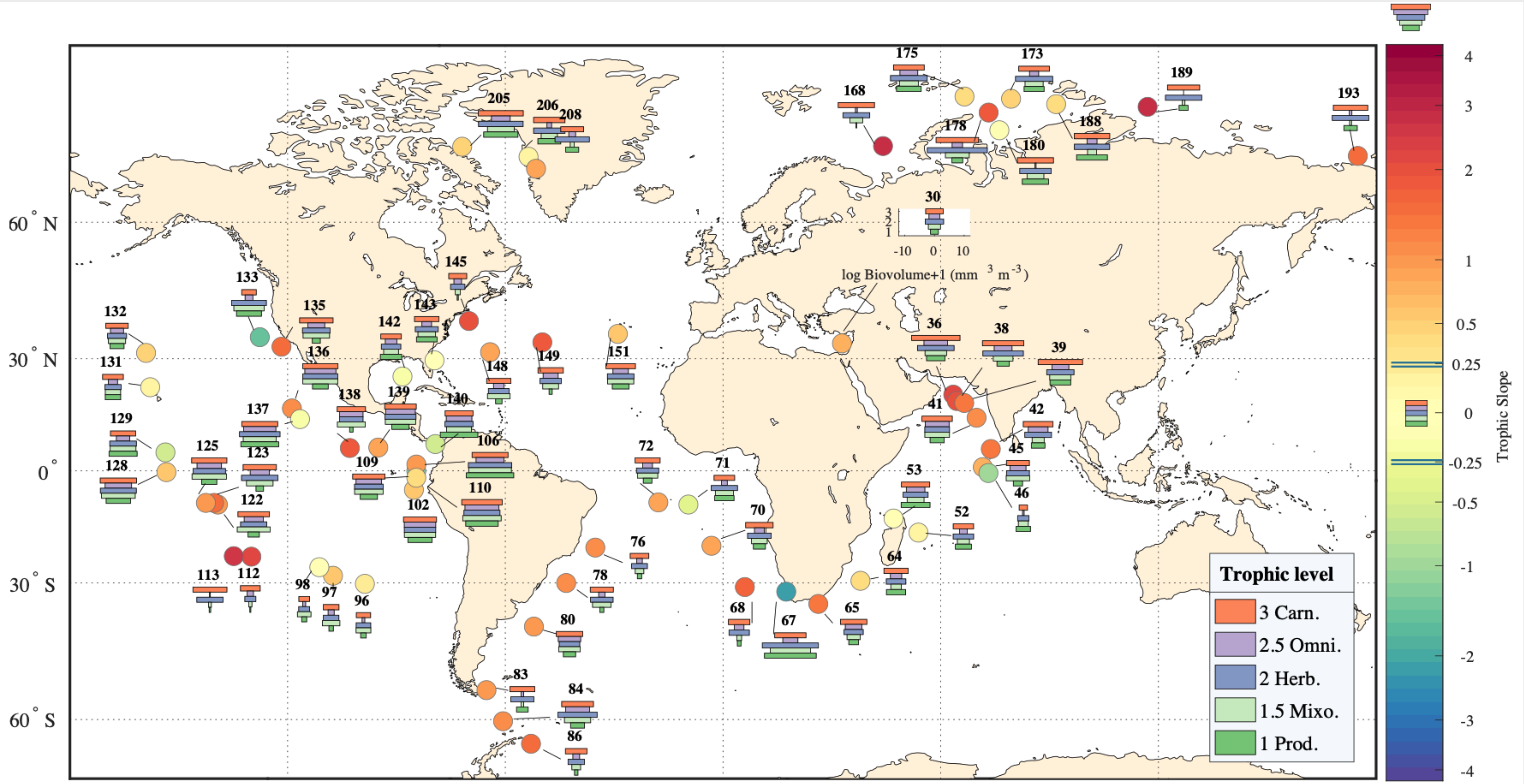


Biard T, Stemmann L, Picheral M, Mayot N, Vandromme P, Hauss H, Gorsky G, Guidi L, Kiko R, Not F (2016) In situ imaging reveals the biomass of giant protists in the global ocean. Nature 532:504.

What for?

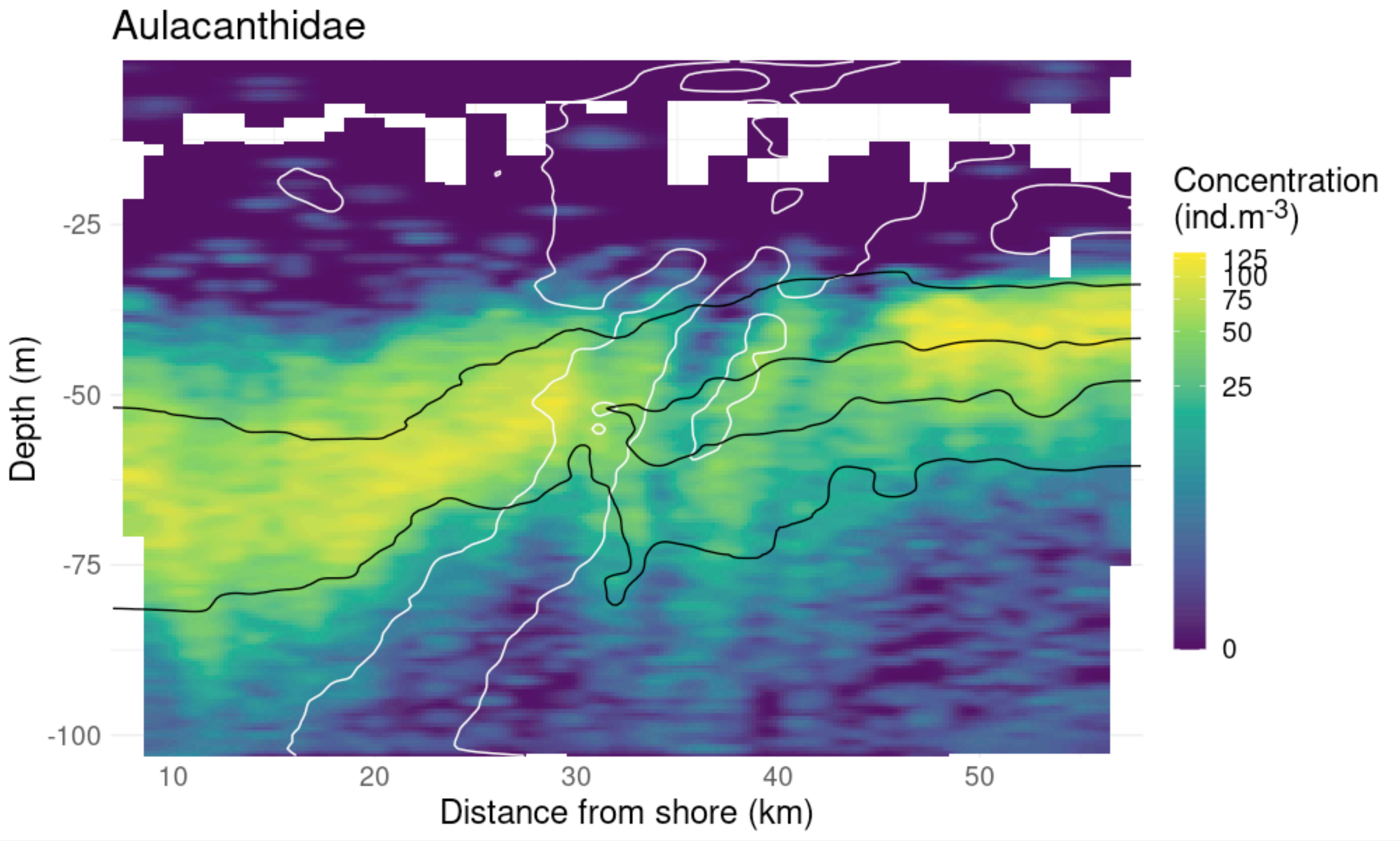
Trophic status of ecosystems

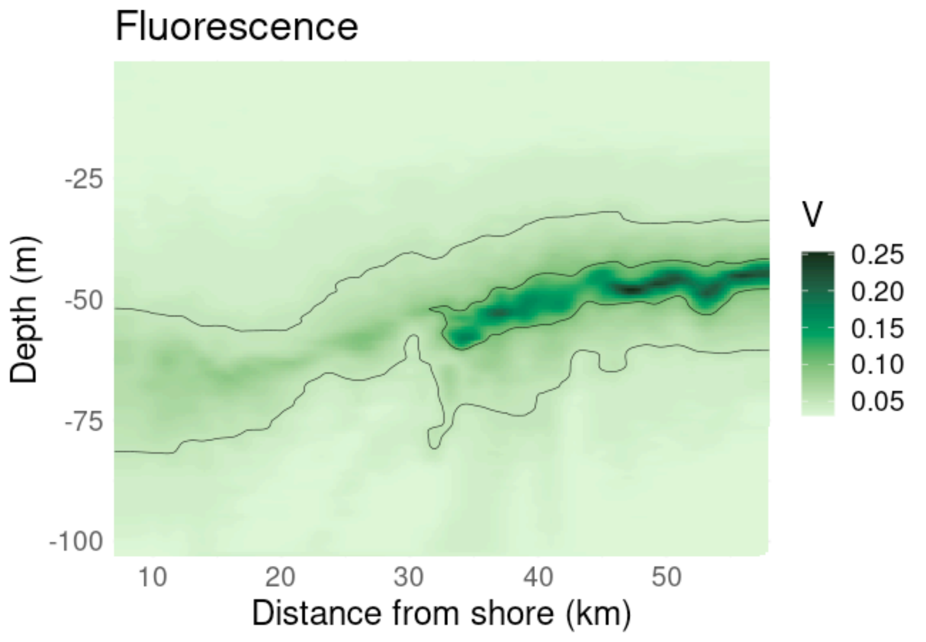


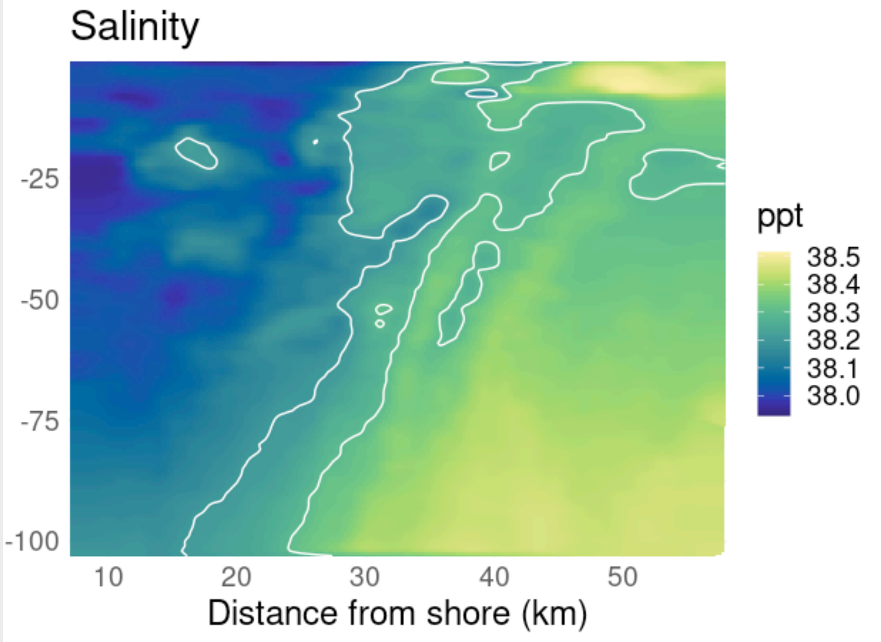


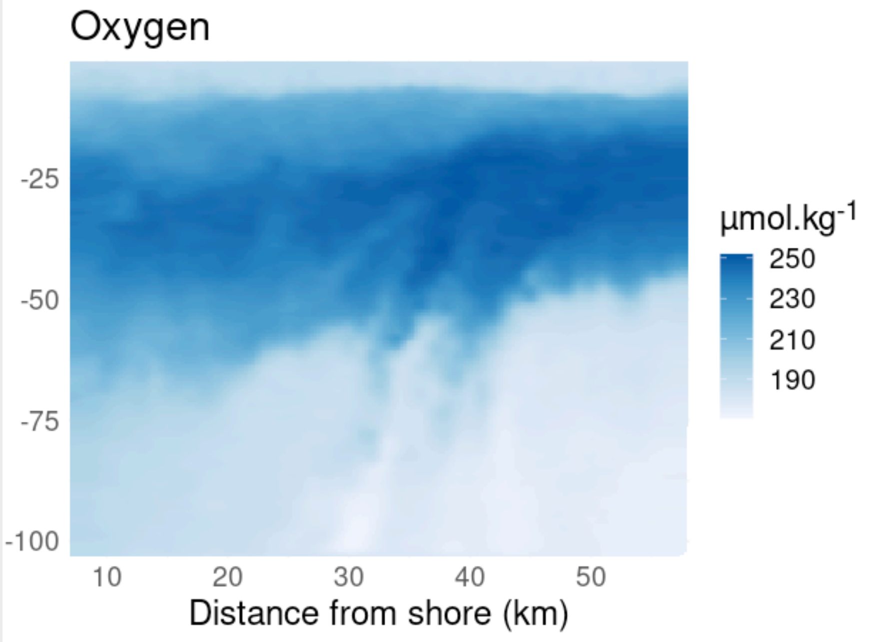
What for?

Super high resolution sampling



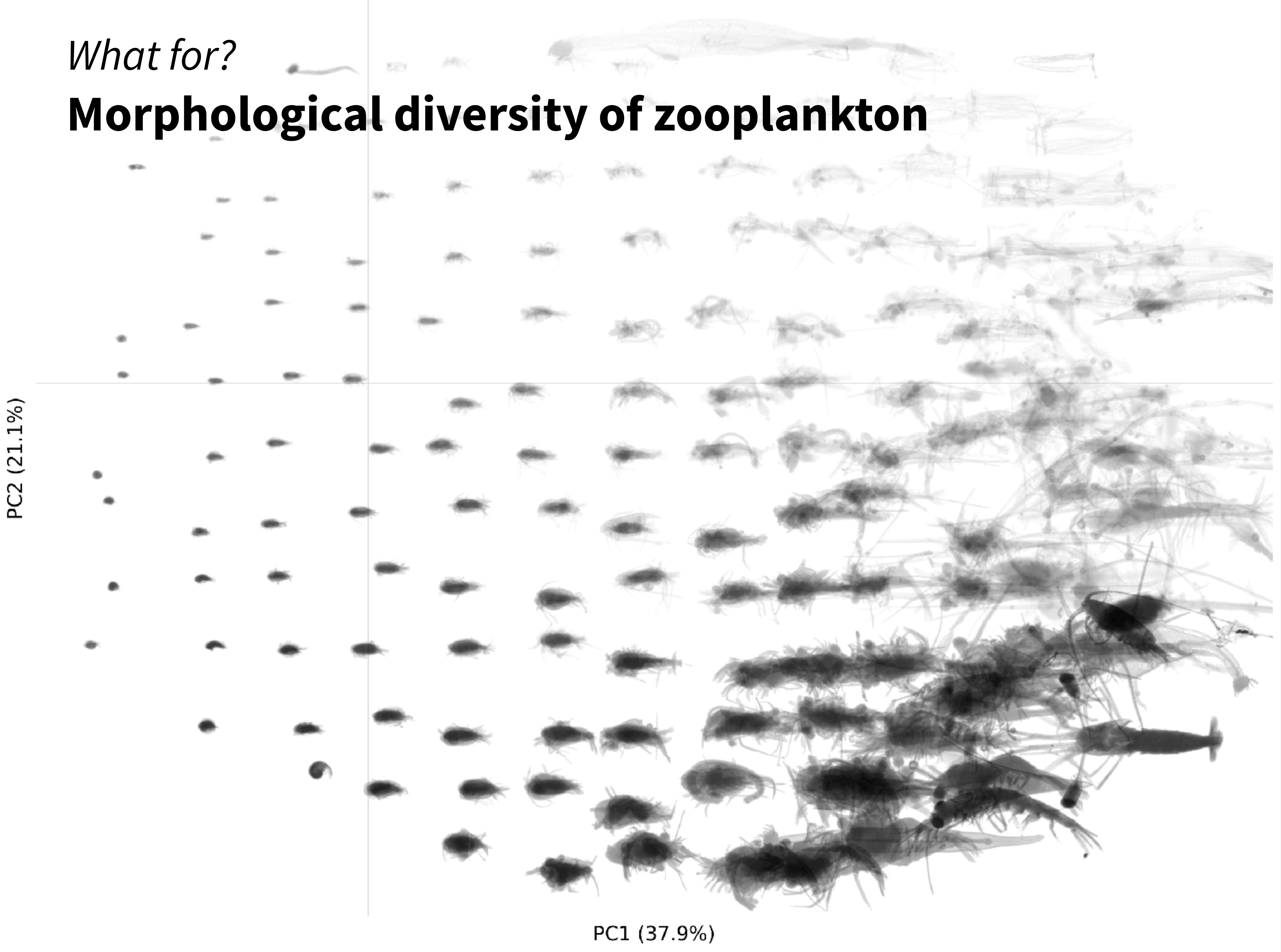






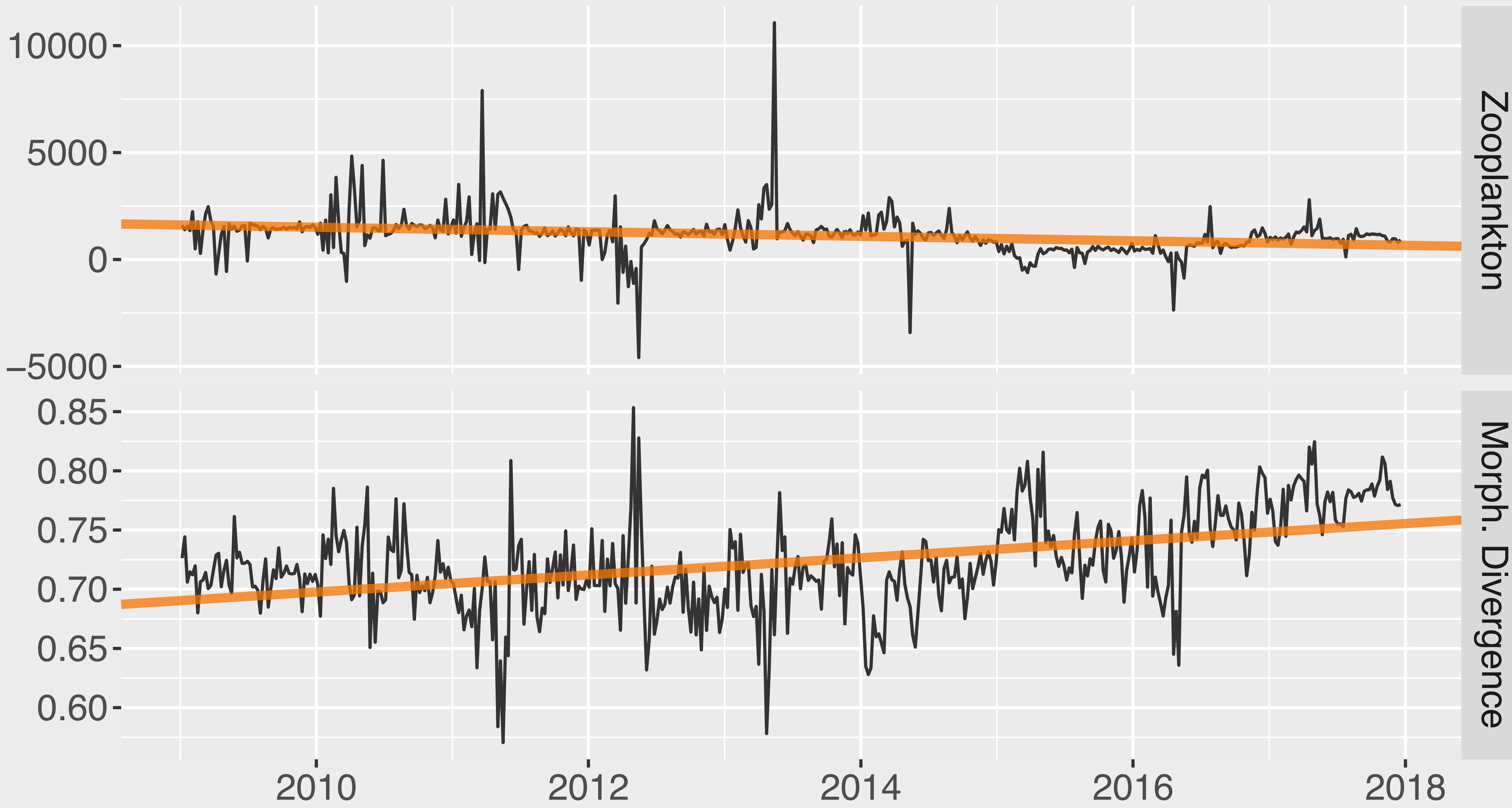
What for?

Morphological diversity of zooplankton



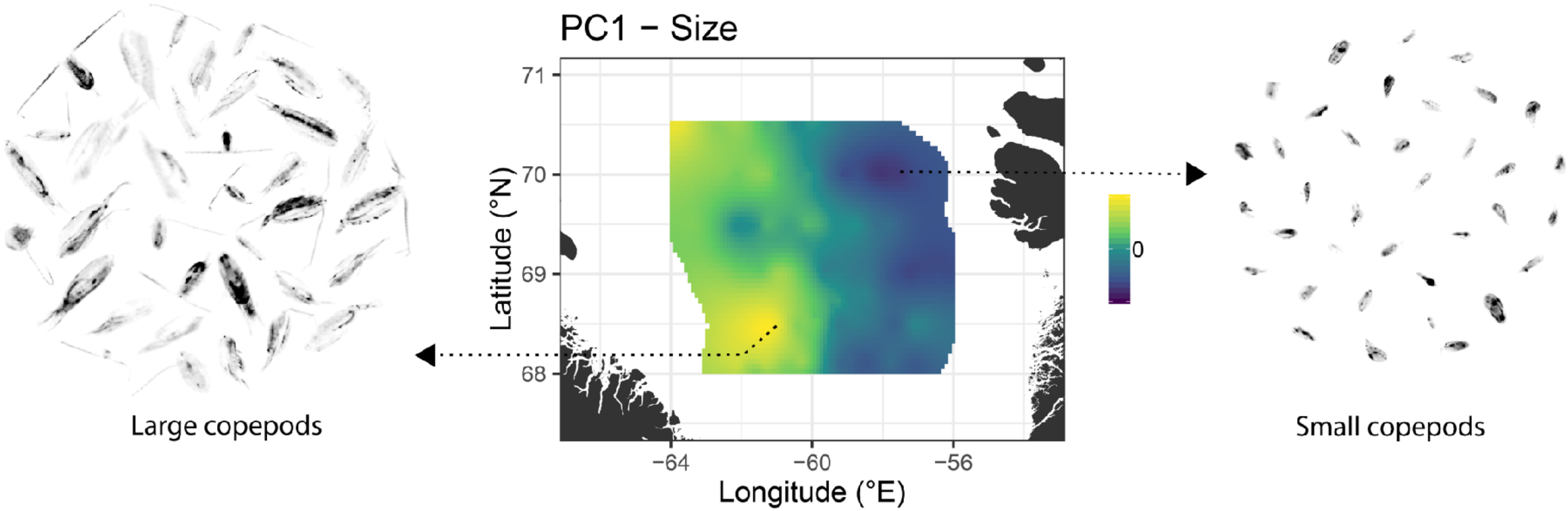
What for?

Morphological diversity of zooplankton



What for?

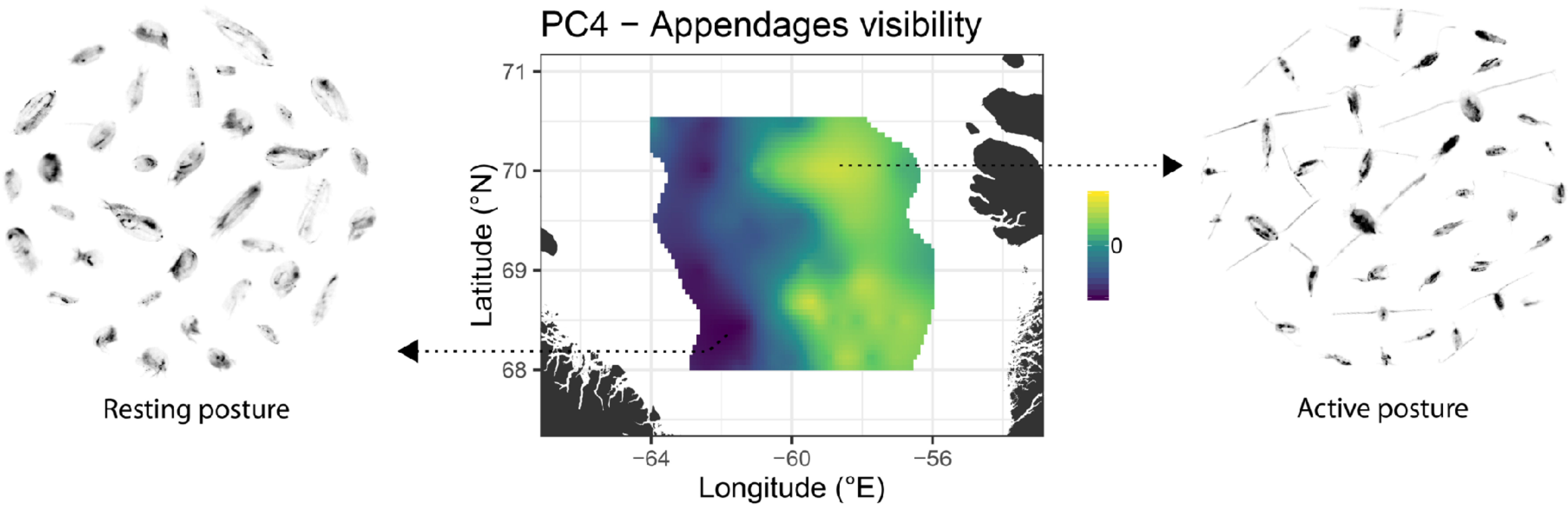
In situ behaviour of organisms



Vilgrain L, Maps F, Picheral M, Babin M, Aubry C, Irisson J-O, Ayata S-D (2021) Trait-based approach using in situ copepod images reveals contrasting ecological patterns across an Arctic ice melt zone. Limnology and Oceanography 66:1155–1167.

What for?

In situ behaviour of organisms



Vilgrain L, Maps F, Picheral M, Babin M, Aubry C, Irisson J-O, Ayata S-D (2021) Trait-based approach using in situ copepod images reveals contrasting ecological patterns across an Arctic ice melt zone. Limnology and Oceanography 66:1155–1167.