



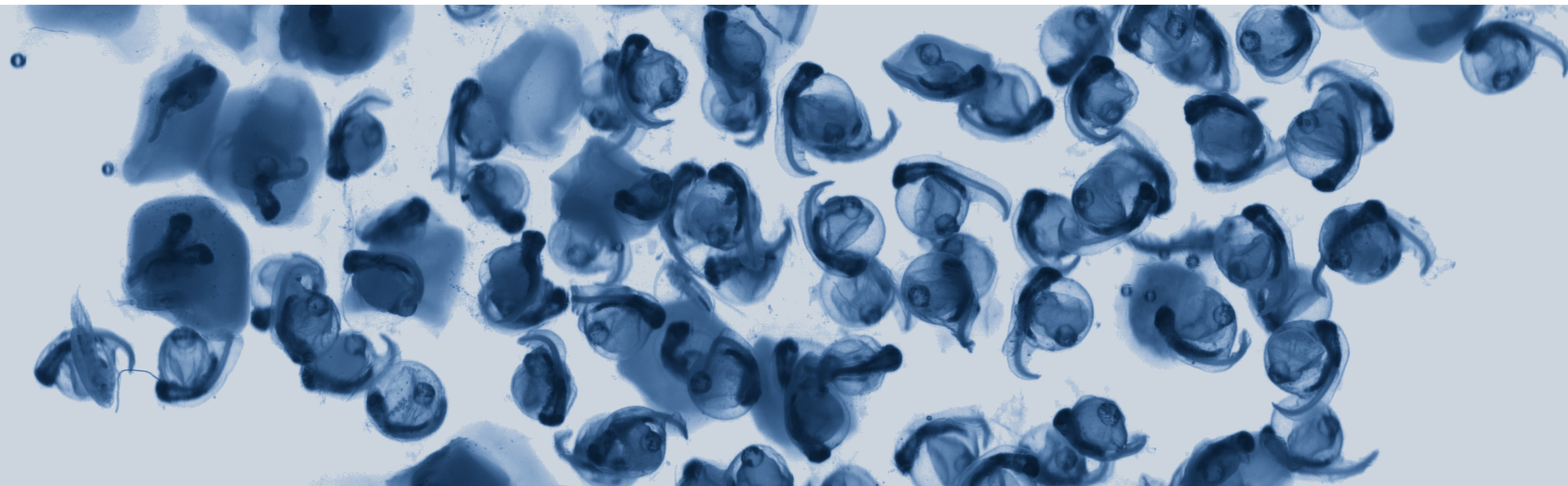
ICES
CIEM



SCIENCES
SORBONNE
UNIVERSITÉ

Increased usage of deep learning systems in marine ecology

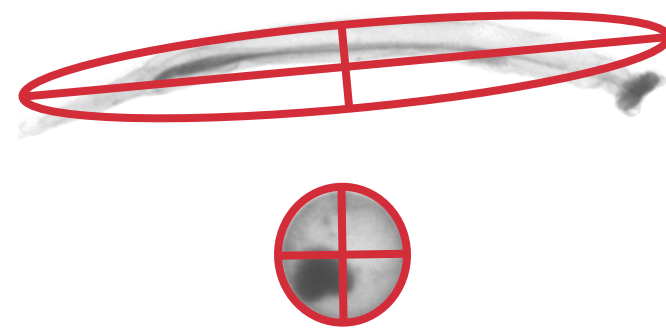
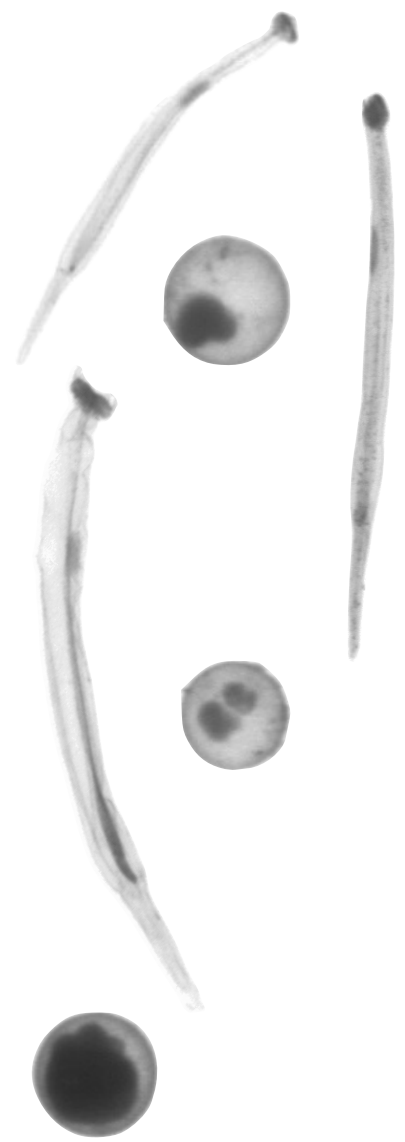
Four years of WGMLEARN



Computer vision vs machine learning vs deep learning

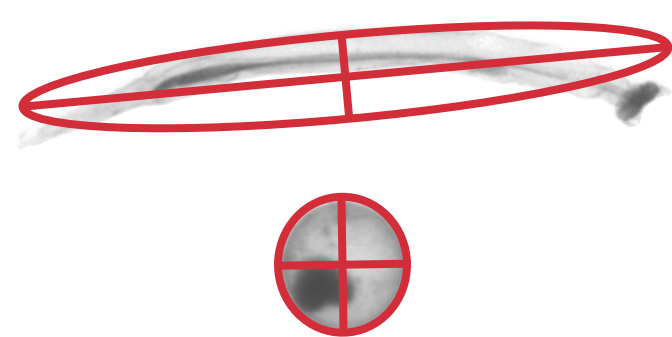


Computer vision vs machine learning vs deep learning

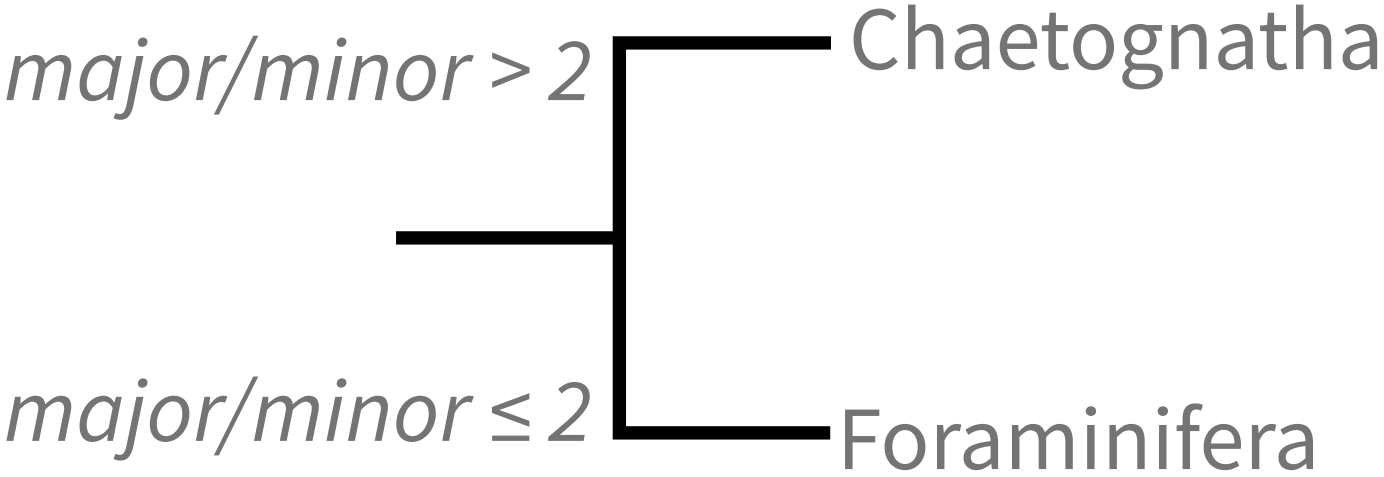


Fit ellipse
↳ Compute major/minor

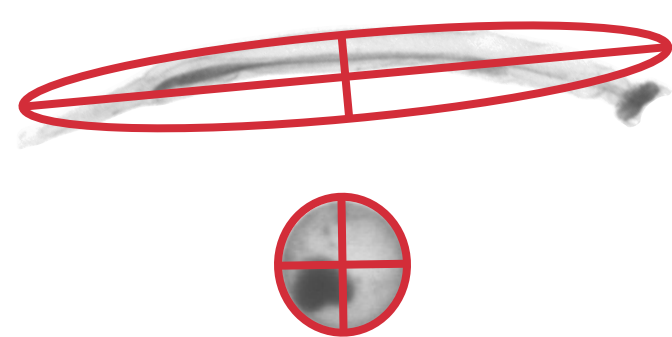
Computer vision vs machine learning vs deep learning



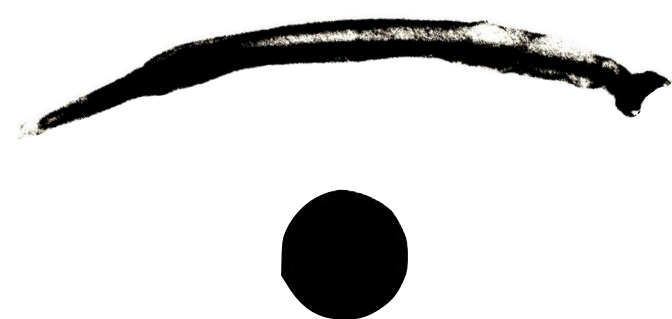
Fit ellipse
↳ Compute major/minor



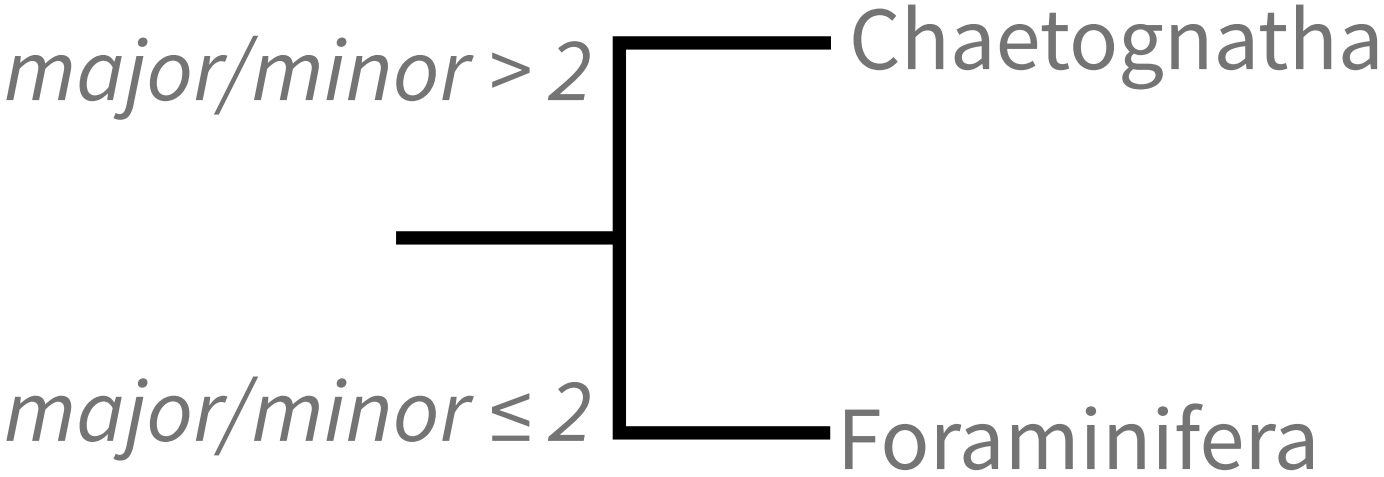
Computer vision vs machine learning vs deep learning



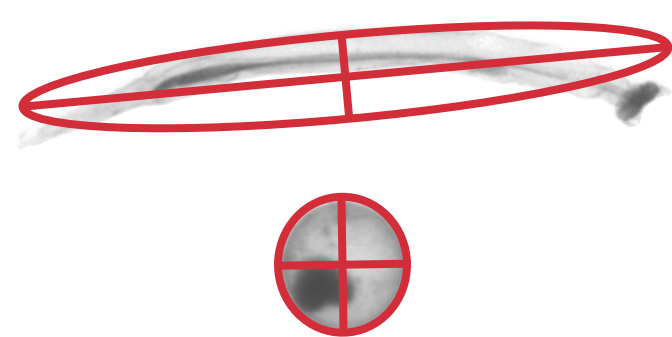
Fit ellipse
↳ Compute major/minor



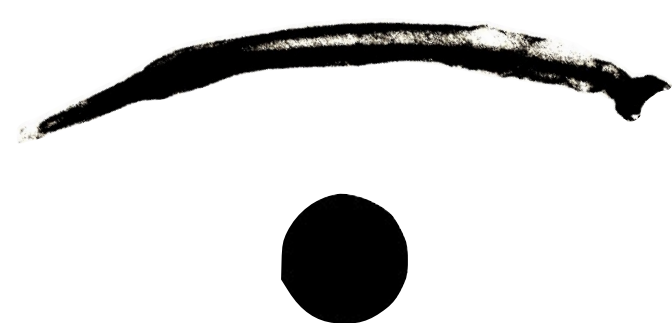
Hu moments
Best fitting ellipse
Grey levels co-occurrence matrix
...



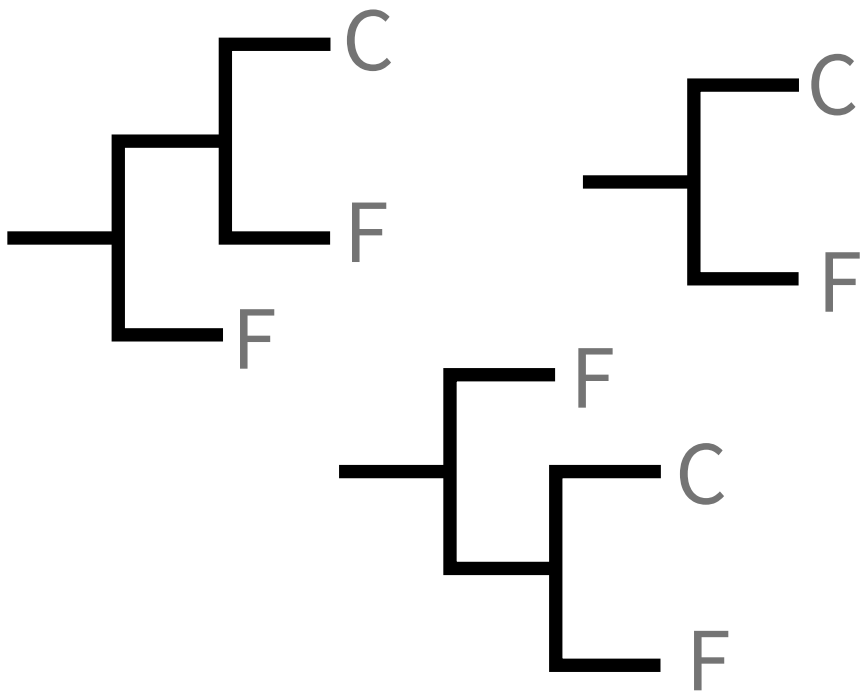
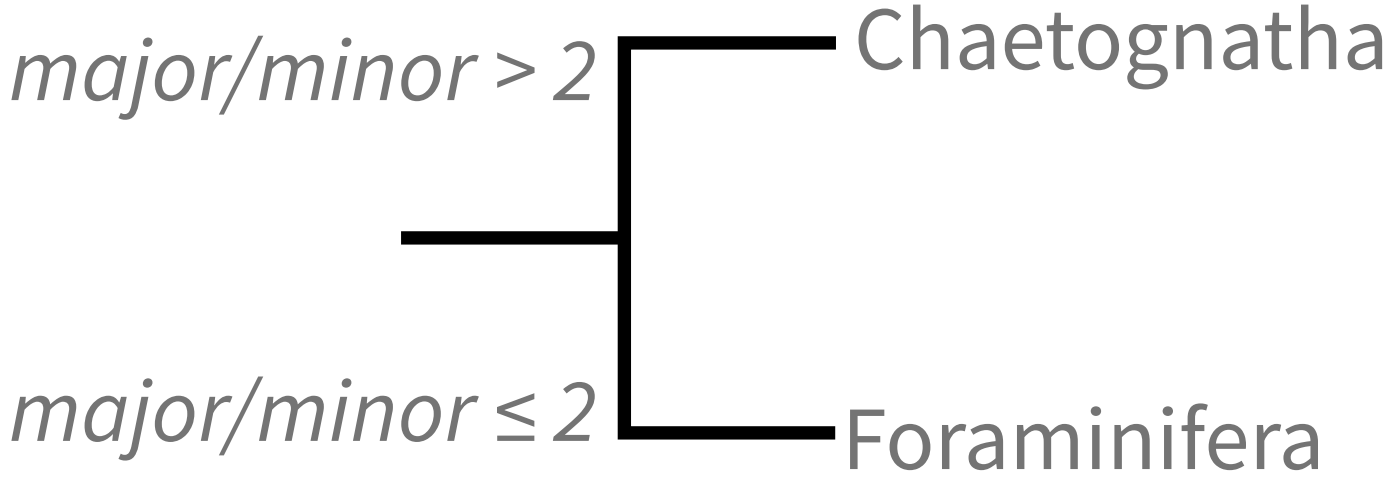
Computer vision vs machine learning vs deep learning



Fit ellipse
↳ Compute major/minor

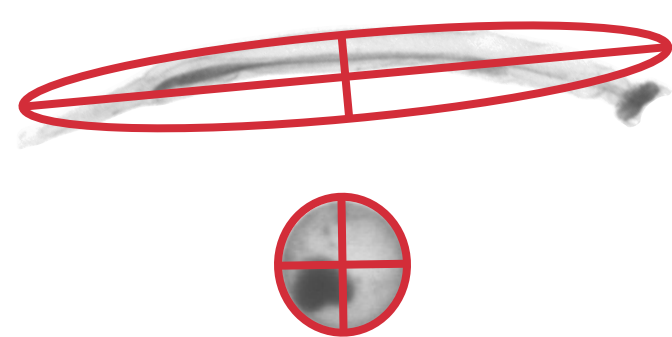


Hu moments
Best fitting ellipse
Grey levels co-occurrence matrix
...

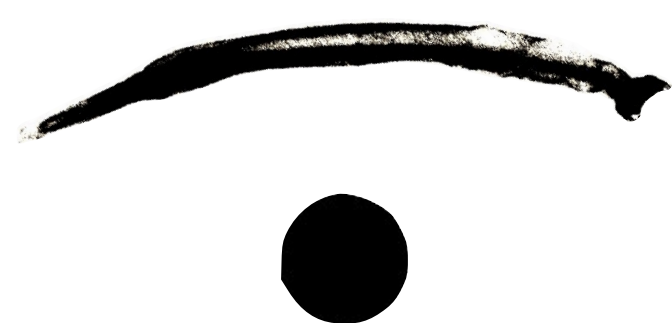


Majority vote
→ Foraminifera

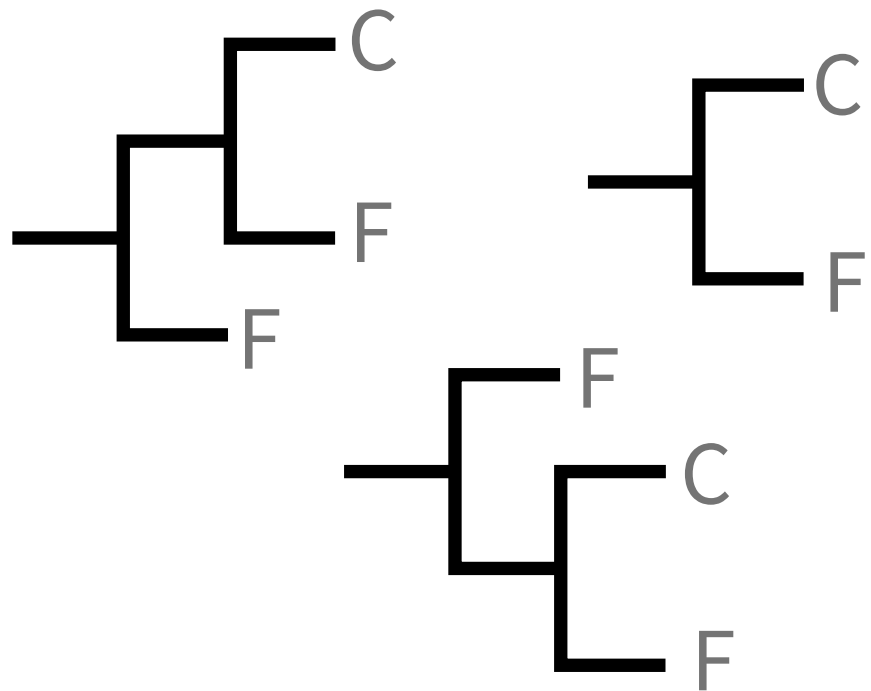
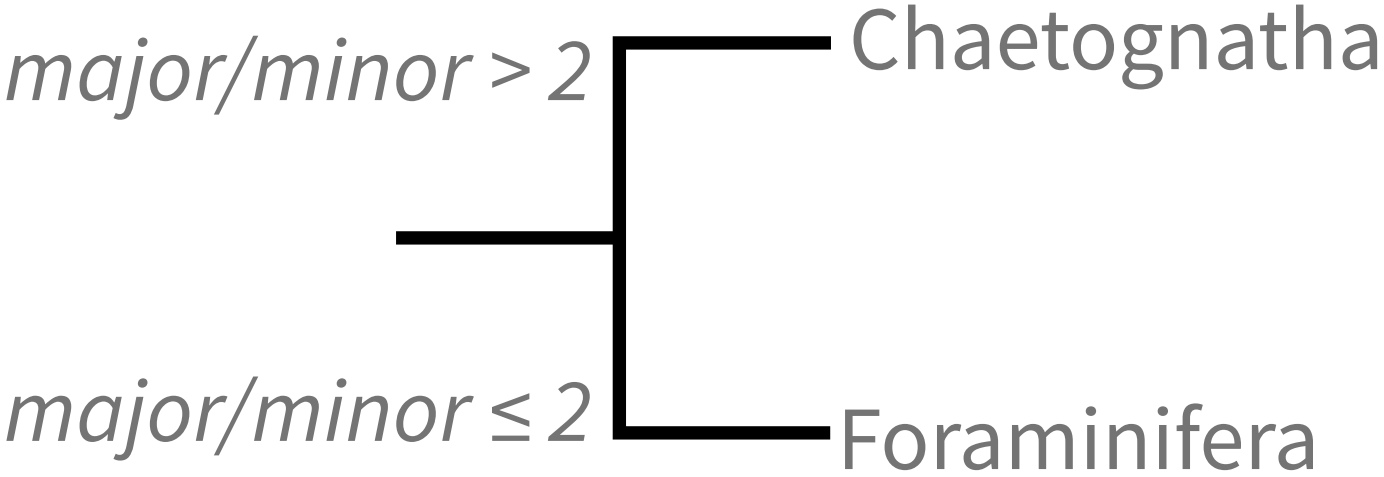
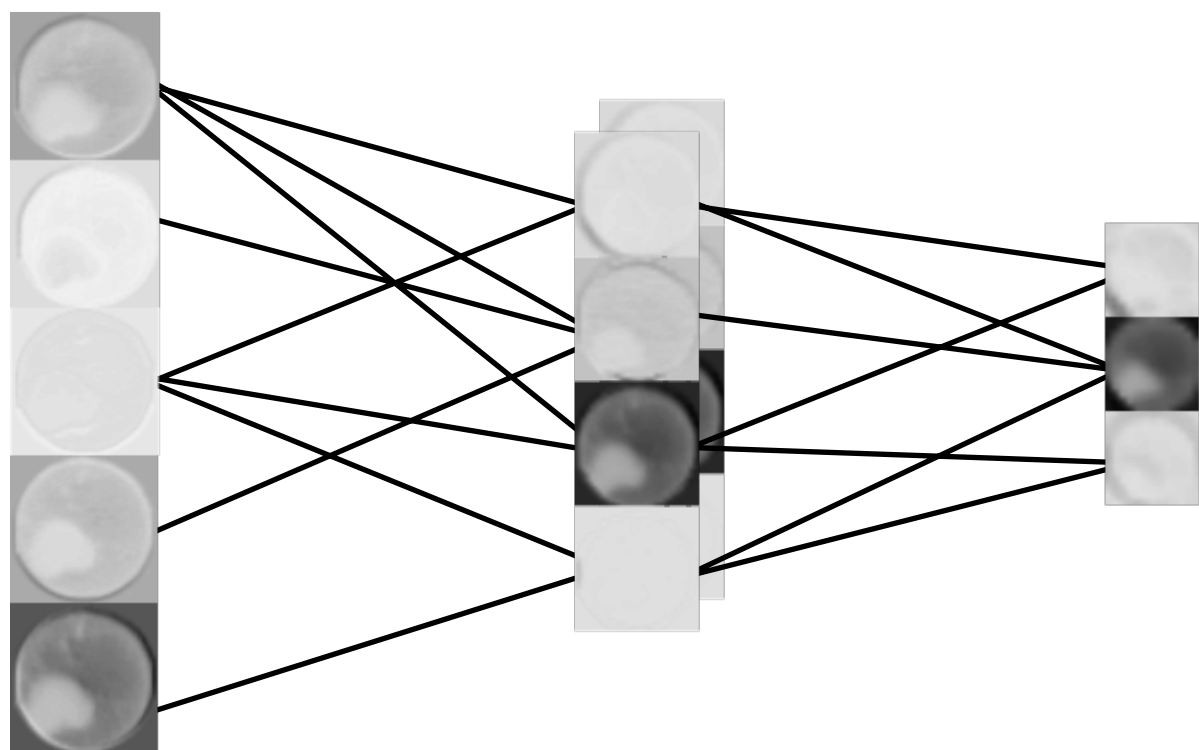
Computer vision vs machine learning vs deep learning



Fit ellipse
↳ Compute major/minor

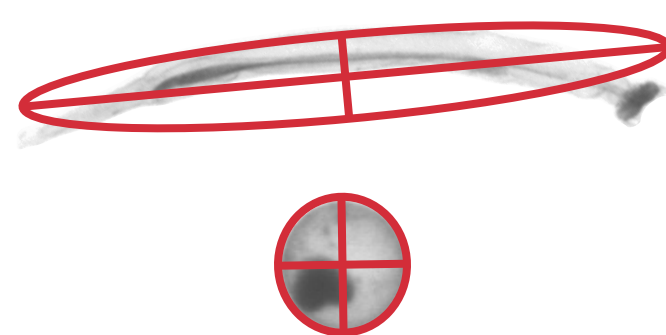


Hu moments
Best fitting ellipse
Grey levels co-occurrence matrix
...

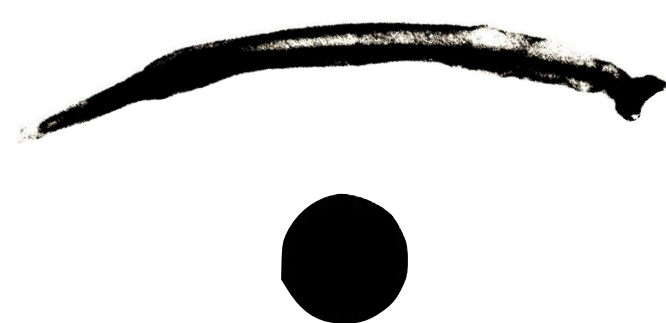


Majority vote → Foraminifera

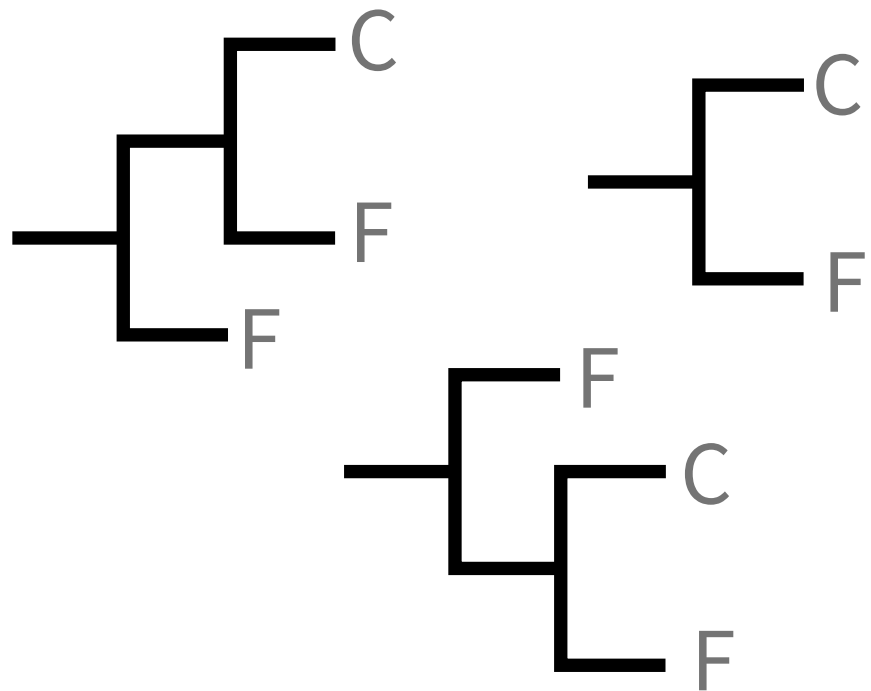
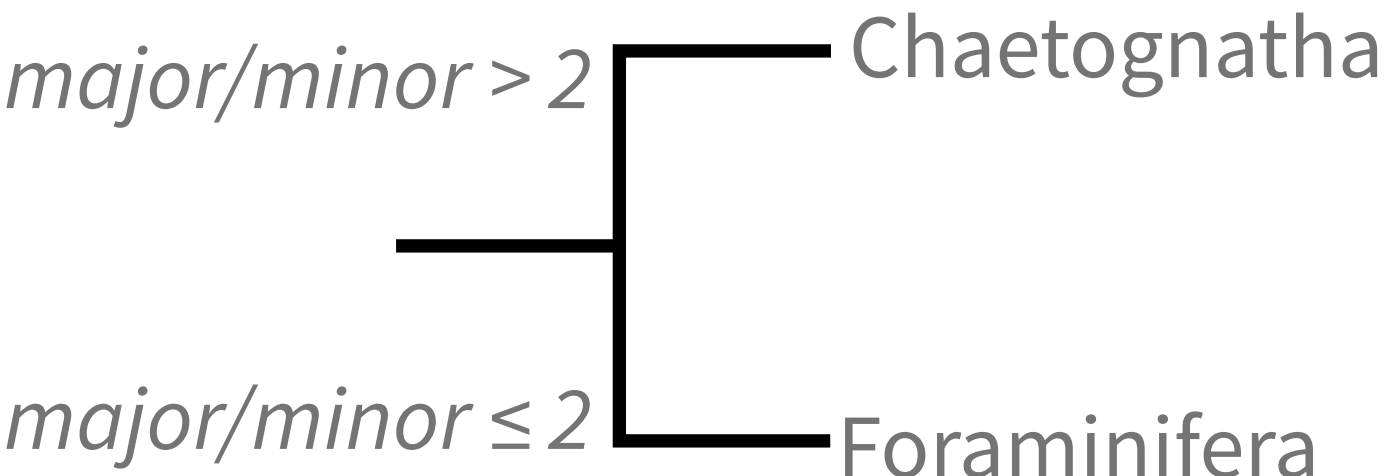
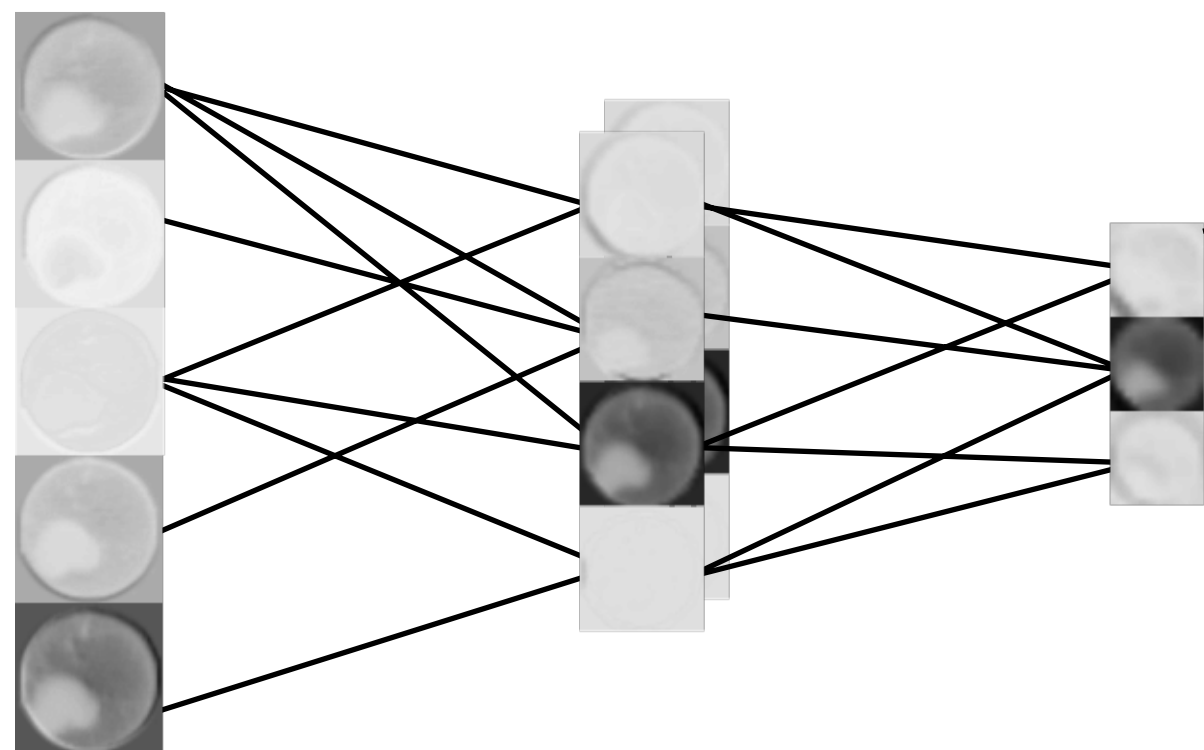
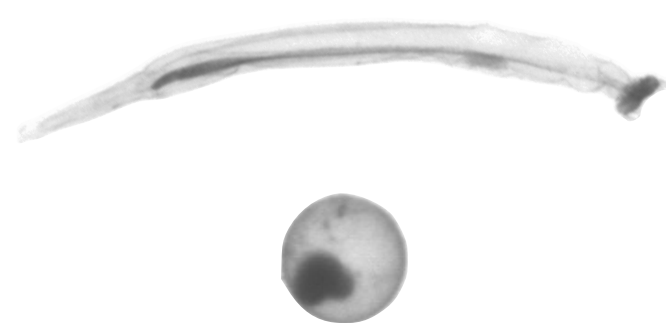
Computer vision vs machine learning vs deep learning



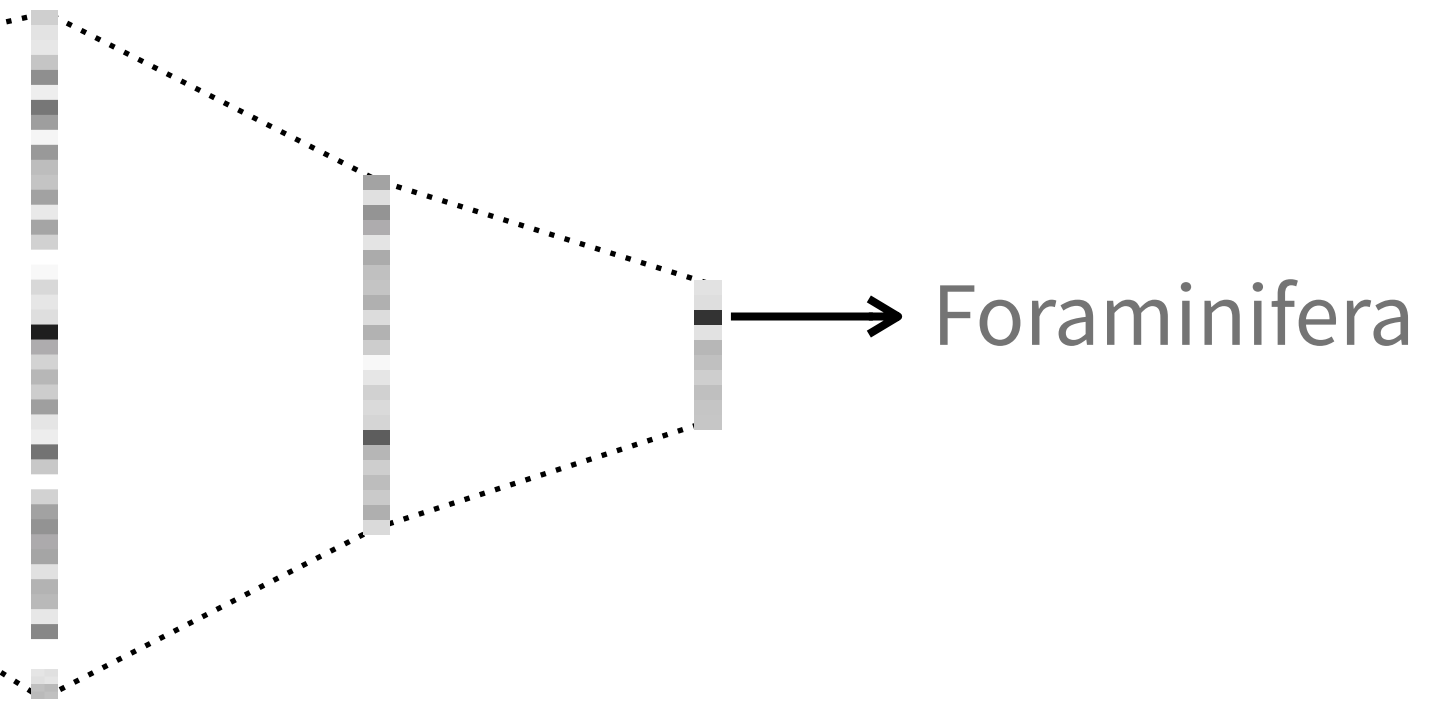
Fit ellipse
↳ Compute major/minor



Hu moments
Best fitting ellipse
Grey levels co-occurence matrix
...

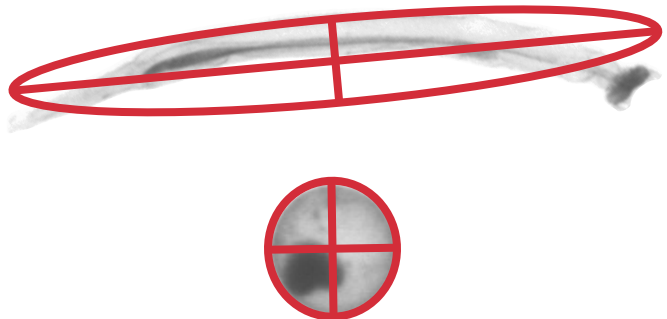
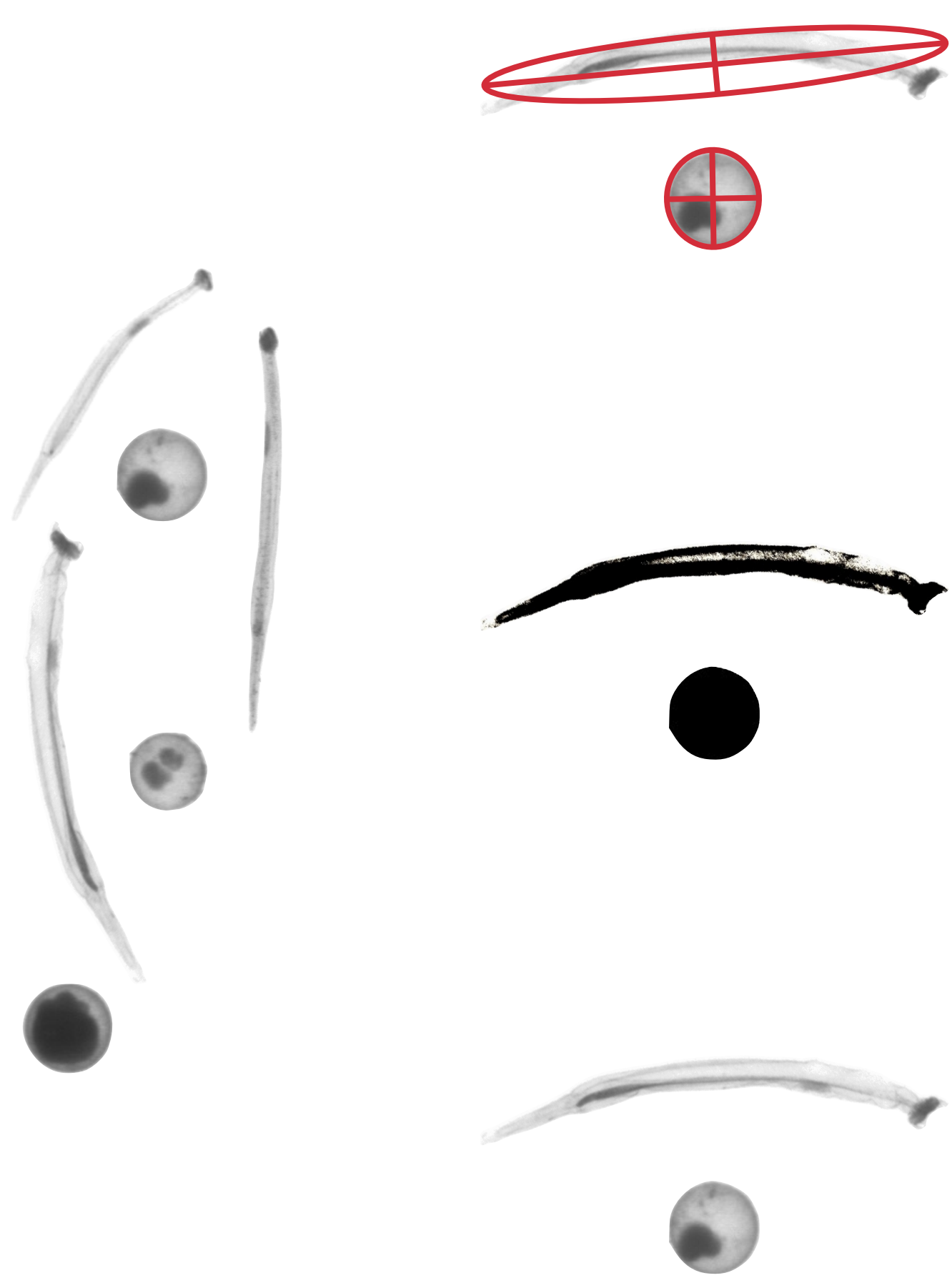


Majority vote
→ Foraminifera

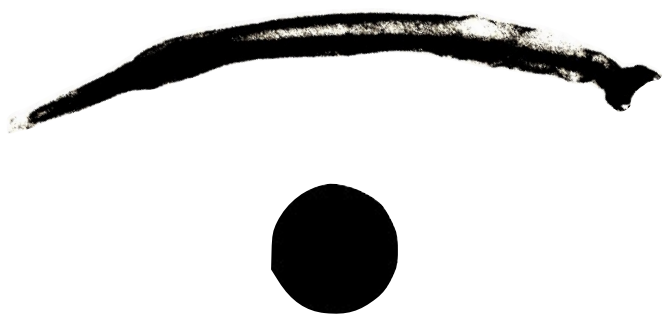


Computer vision vs machine learning vs deep learning

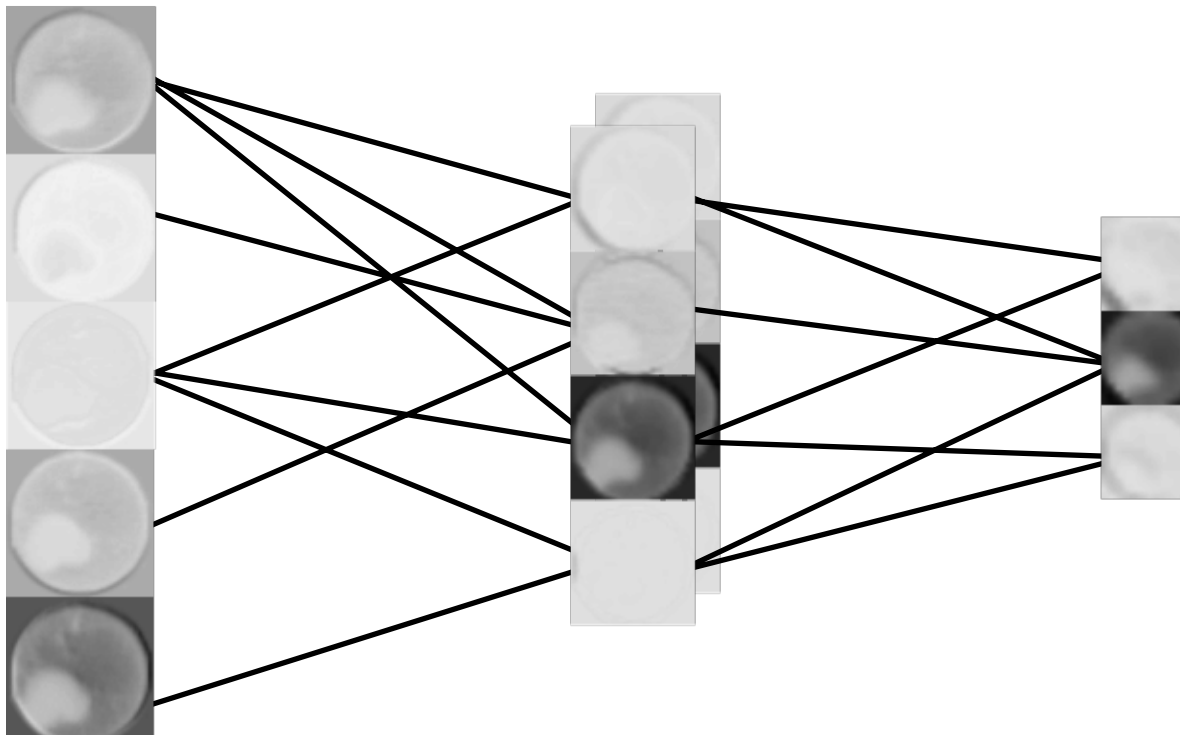
Feature extraction



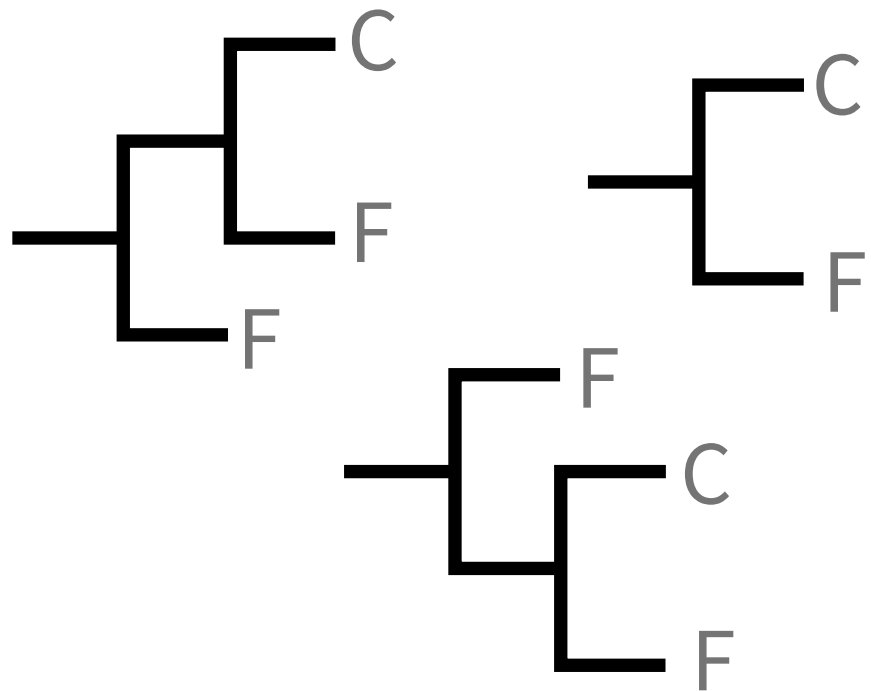
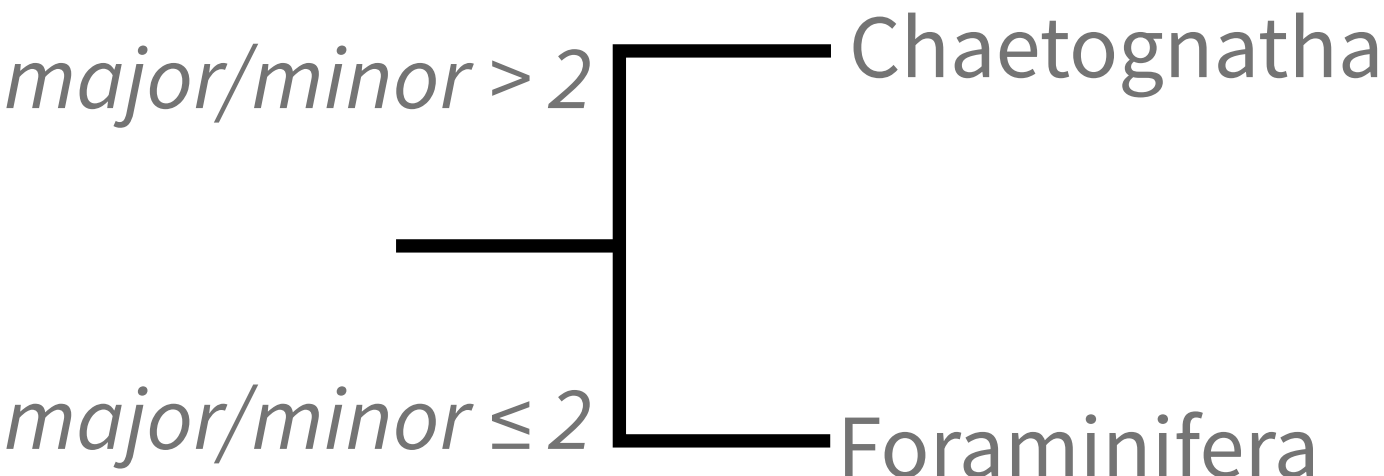
Fit ellipse
↳ Compute major/minor



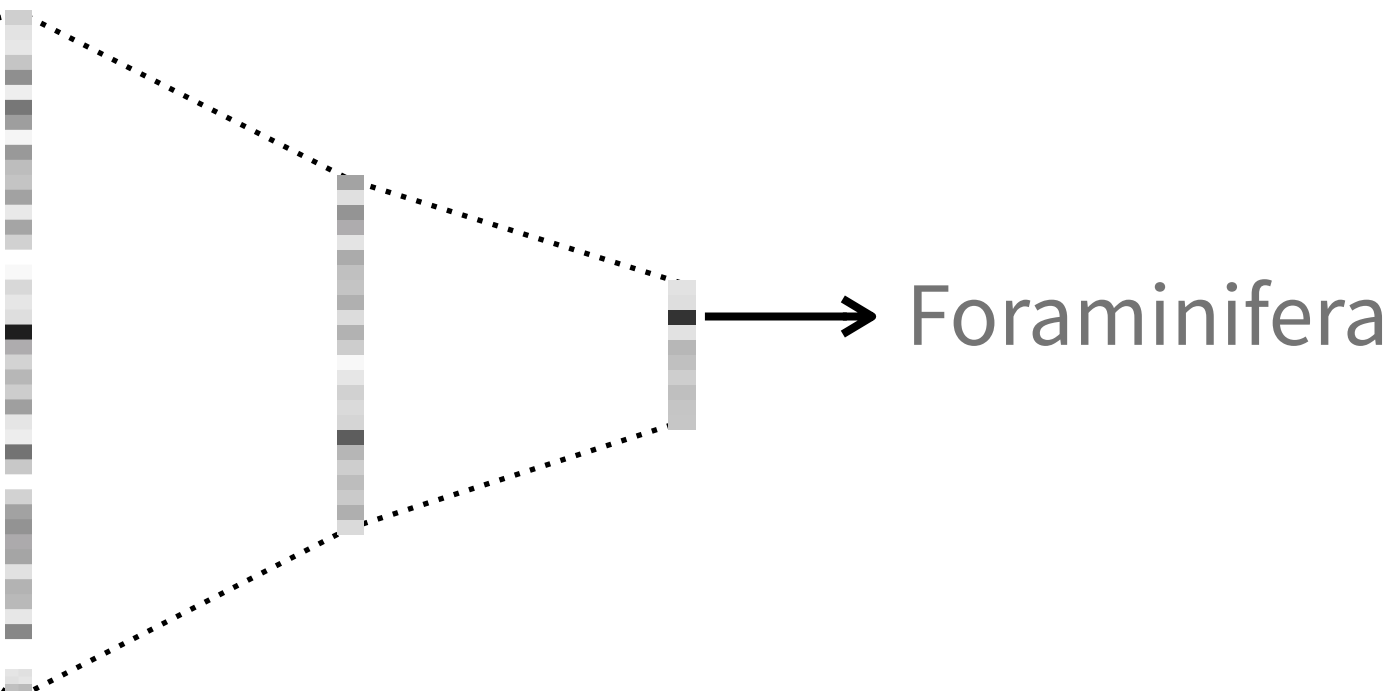
Hu moments
Best fitting ellipse
Grey levels co-occurrence matrix
...



Classification



Majority vote
→ Foraminifera



WGMLEARN’s study

Assemble potential literature on ML and marine ecology (through search terms + bibliography mining)

Curate to keep peer reviewed, using ML, effectively on marine

Active for target classif

Recently added

Peer reviewed

Recent

My Publications

Duplicate Items

Unfiled Items

Bin

Group Libraries

complex

wgmlearn

_Guided tour (group C paper)

_HOW TO

_Other

_Related work (e.g. ML but not...)

_Reviews/prospectives of mari...

Acoustics

Active for seabed/sediment ...

Active for target classif

Passive

Geolocation (VMS, AIS, etc.)

Gridded data

In situ

Satellite

Image/video

Benthos

Litter

Nekton

Plankton

method:Autoencoder

method:Bayesian–Network

method:Classification–And–Regressi...

method:Convolutional–Neural–Netw...

method:Discriminant–Analysis

method:Ensemble

method:Generalised–Additive–Model

method:Gradient–Boosted–Trees

method:instance–based–learning

method:iso–cluster

method:k–means

QTitle, Creator, Year

Title	Creator	Publication	Year	E...	
▶ Semi-supervised target classification in multi-frequency echosounder data	Choi et al.	ICES Journal of ...	2021	2 ...	
▶ Machine learning to detect bycatch risk: Novel application to echosounder buoy...	Mannoc...	Biological Cons...	2021	9 ...	
▶ Differentiation of two swim bladdered fish species using next generation wideb...	Gugele ...	Scientific Reports	2021	2 ...	
▶ Detecting Marine Species in Echograms via Traditional, Hybrid, and Deep Learni...	Marque...	2020 25th Inter...	2021	1 ...	
▶ Deep Learning for Deep Waters: An Expert-in-the-Loop Machine Learning Fram...	Ryazan...	Journal of Marin...	2021	3 ...	
▶ Machine learning for characterizing tropical tuna aggregations under Drifting Fi...	Baidai e...	Fisheries Resear...	2020	14...	
▶ EchoBERT: A Transformer-Based Approach for Behavior Detection in Echograms	Maloy	IEEE Access	2020	4 ...	
▶ Complex data labeling with deep learning methods: Lessons from fisheries aco...	Sarr et al.	ISA Transactions	2020	7 ...	
▶ Compact representation of temporal processes in echosounder time series via ...	Lee and...	The Journal of t...	2020	4 ...	
▶ Automated classification of schools of the silver cyprinid Rastrineobola argente...	Proud e...	ICES Journal of ...	2020	8 ...	
▶ Assessing the Repeatability of Automated Seafloor Classification Algorithms, wi...	Zelada L...	Remote Sensing	2020	20...	
▶ Acoustic classification in multifrequency echosounder data using deep convolut...	Brautas...	ICES Journal of ...	2020	19...	
▶ 3D Fine-scale Terrain Variables from Underwater Photogrammetry: A New Appr...	Prado et...	Remote Sensing	2020	10...	
▶ Identifying small pelagic Mediterranean fish schools from acoustic and environ...	Aronica ...	Ecological Infor...	2019	10...	
▶ Development of benthic monitoring approaches for salmon aquaculture sites u...	Armstro...		2019		
▶ Counting and sizing Atlantic bluefin tuna schools using medium range sonars ...	Uranga ...	Continental She...	2019	3 ...	
▶ Study on Echo Features and Classification Methods of Fish Species	Shang a...	2018 10th Inter...	2018	4 ...	
▶ Robust clustering methodology for multi-frequency acoustic data: A review of s...	Peña	Fisheries Resear...	2018	9 ...	
▶ Acoustic target classification	Korneliu...	ICES Cooperativ...	2018		
▶ Discriminating fish species by an Echo sounder in a set-net using a CNN	Hirama ...	2017 21st Asia ...	2017	8 ...	
▶ Detecting the presence-absence of bluefin tuna by automated analysis of medi...	Uranga ...	PLOS ONE	2017	13...	

47 items in this view

WGMLEARN's study

Assemble potential literature on ML and marine ecology (through search terms + bibliography mining)

Curate to keep peer reviewed, using ML, effectively on marine

Tag papers according to data, task, method, or other aspects (meta)

General rules

- Do not use spaces, use - instead.
- Avoid tags that are too specific: the tag should be applicable to multiple (>10) items to be useful.

For the following categories, use category:tag

category	n	task	n	method	n	meta	n
data:acoustics	50	task:supervised	72	method:Artificial-Neural-Network	100+	meta:code-available	24
data:active-acoustics	69	task:classification	24	method:autoencoder	2	meta:data-available	31
data:passive-acoustics	100+	task:binary-classification	40	method:bagging	3	meta:dataset	5
		task:multiclass-classification	76	method:bayesian-classifiers	15	meta:key-paper	34
		task:multilabel-classification		method:bayesian-networks	20	meta:key-paper-(maybe)	1
data:geolocation	49			method:bayesian-regression	1	meta:review	10
data:AIS	1	task:regression	2	method:C4.5	1		
data:VMS	3	task:single-output-regression	42	method:canonical-correlation-forest	1		
		task:multi-output-regression	10	method:Classification-And-Regression-Trees	39		
				method:co-occurrence-network	1		
		task:semi-supervised	3	method:collaborative-conditional-random-field	1		
data:remote-sensing	100+	task:few-shot-learning		method:competitive-learning	1		
data:vertical-profiles		task:zero-shot-learning		method:Convolutional-Neural-Network	100+		
data:radiometry	42			method:decision-trees	18		
				method:deep-belief-network	5		
data:image	8	task:unsupervised	23	method:ensemble	3		
data:video	36	task:clustering	40	method:extremely-randomised-trees	1		
		task:dimensionality-reduction	5	method:fuzzy-logic	9		
data:omics	11			method:fuzzy-rule-based-bayesian-inference	1		
		task:representation-learning	4	method:gaussian-mixture	19		
		task:feature-extraction	37	method:gaussian-process	2		
data:flow-cytometry	23	task:feature-selection	4	method:Generalised-Additive-Model	4		
data:raman-spectroscopy	3	task:variable-selection	2	method:Generalised-Linear-Model	3		
data:mass-spectrometry	6			method:Generative-Adversarial-Network	1		
		task:object-detection		method:genetic-algorithm	10		
data:data-series	20	task:object-segmentation	8	method:Gradient-Boosted-Trees	71		
data:aerial-image	2			method:gradient-forest	1		
		task:outlier-detection	1	method:Hidden-Markov-Model	4		
		task:pattern-recognition	1	method:hierarchical-clustering	3		
				method:instance-based-learner	1		

WGMLEARN’s study

Assemble potential literature on ML and marine ecology (through search terms + bibliography mining)

Curate to keep peer reviewed, using ML, effectively on marine

Tag papers according to data, task, method, or other aspects (meta)

Select papers to cite

Write summary in sections organised according to input data

Machine learning in marine ecology: an overview of techniques and applications

Peter Rubbens^{1,2}, Stephanie Brodie³, Tristan Cordier^{4,5}, Diogo Destro Barcellos⁶, Paul Devos⁷, Jose A. Fernandes-Salvador⁸, Jennifer I Fincham ⁹, Alessandra Gomes⁶, Nils Olav Handegard ¹⁰, Kerry Howell¹¹, Cédric Jamet¹², Kyrre Helda Kartveit¹⁰, Hassan Moustahfid¹³, Clea Parcerisas ^{1,7}, Dimitris Politikos ¹⁴, Raphaëlle Sauzède ¹⁵, Maria Sokolova ¹⁶, Laura Uusitalo ^{17,18}, Laure Van den Bulcke^{19,20}, Aloysius T. M. van Helmond²¹, Jordan T. Watson ^{22,23}, Heather Welch³, Oscar Beltran-Perez ²⁴, Samuel Chaffron^{25,26}, David S. Greenberg²⁷, Bernhard Kühn²⁸, Rainer Kiko ^{29,30}, Madiop Lo³¹, Rubens M. Lopes ⁶, Klas Ove Möller ³², William Michaels³³, Ahmet Pala^{34,10}, Jean-Baptiste Romagnan³⁵, Pia Schuchert³⁶, Vahid Seydi³⁷, Sebastian Villasante³⁸, Ketil Malde ^{10,39}, and Jean-Olivier Irisson ^{29,*}

¹Flanders Marine Institute (VLIZ), 8400 Oostende, Belgium

²Kytos BV, Technologiepark-Zwijnaarde 82, 9052 Gent, Belgium

³Institute of Marine Science, University of California Santa Cruz, Santa Cruz, CA 95064, USA

⁴Department of Genetics and Evolution, University of Geneva, 1205 Geneva, Switzerland

⁵NORCE Climate, NORCE Norwegian Research Centre AS, Bjerknes Centre for Climate Research, Jahnebakken 5, 5007 Bergen, Norway

⁶Oceanographic Institute, University of São Paulo, Praça do Oceanográfico, 191, 05508-120, São Paulo, Brazil

⁷Department of Information Technology, Research group WAVES, Ghent University, Tech Lane Ghent Science Park, 126, B-9058 Gent, Belgium

⁸AZTI, Marine Research, Basque Research and Technology Alliance (BRTA). Txatxarramendi Ugarte z/g, 48395 Sukarrieta, Spain

⁹Cefas, Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK

¹⁰Institute of Marine Research, Nykirkekaia 1, 5005 Bergen, Norway

¹¹School of Biological and Marine Sciences, University of Plymouth, Drake Circus, Plymouth PL4 8AA, UK

¹²Université du Littoral Côte d'Opale, CNRS, Univ. Lille, IRD, UMR 8187, LOG, Laboratoire d'Océanologie et de Géosciences, F-62930 Wimereux, France

¹³National Oceanic and Atmospheric Administration, US Integrated Ocean Observing System, Silver Spring, MD 20910, USA

¹⁴Institute of Marine Biological Resources and Inland, Hellenic Centre for Marine Research, 16452 Argyroupoli, Greece

¹⁵Sorbonne Université, CNRS, Institut de la Mer de Villefranche, FR3761, F-06230 Villefranche-Sur-Mer, France

¹⁶Wageningen University and Research, Droevendaalsesteeg 1, Building 107, 6708 PB Wageningen, The Netherlands

¹⁷Finnish Environment Institute, Latokartanonkaari 11, FI-00790 Helsinki, Finland

¹⁸Natural Resources Institute Finland (Luke), Latokartanonkaari 9, FI-00790 Helsinki, Finland

¹⁹Flanders Research Institute for Agriculture, Fisheries and Food, Marine Research, Jacobsenstraat 1, 8400 Ostend, Belgium

²⁰Department of Data Analysis and Mathematical Modelling—Knowledge-based Systems Research Group, University of Ghent, Coupure Links 653, 9000 Gent, Belgium

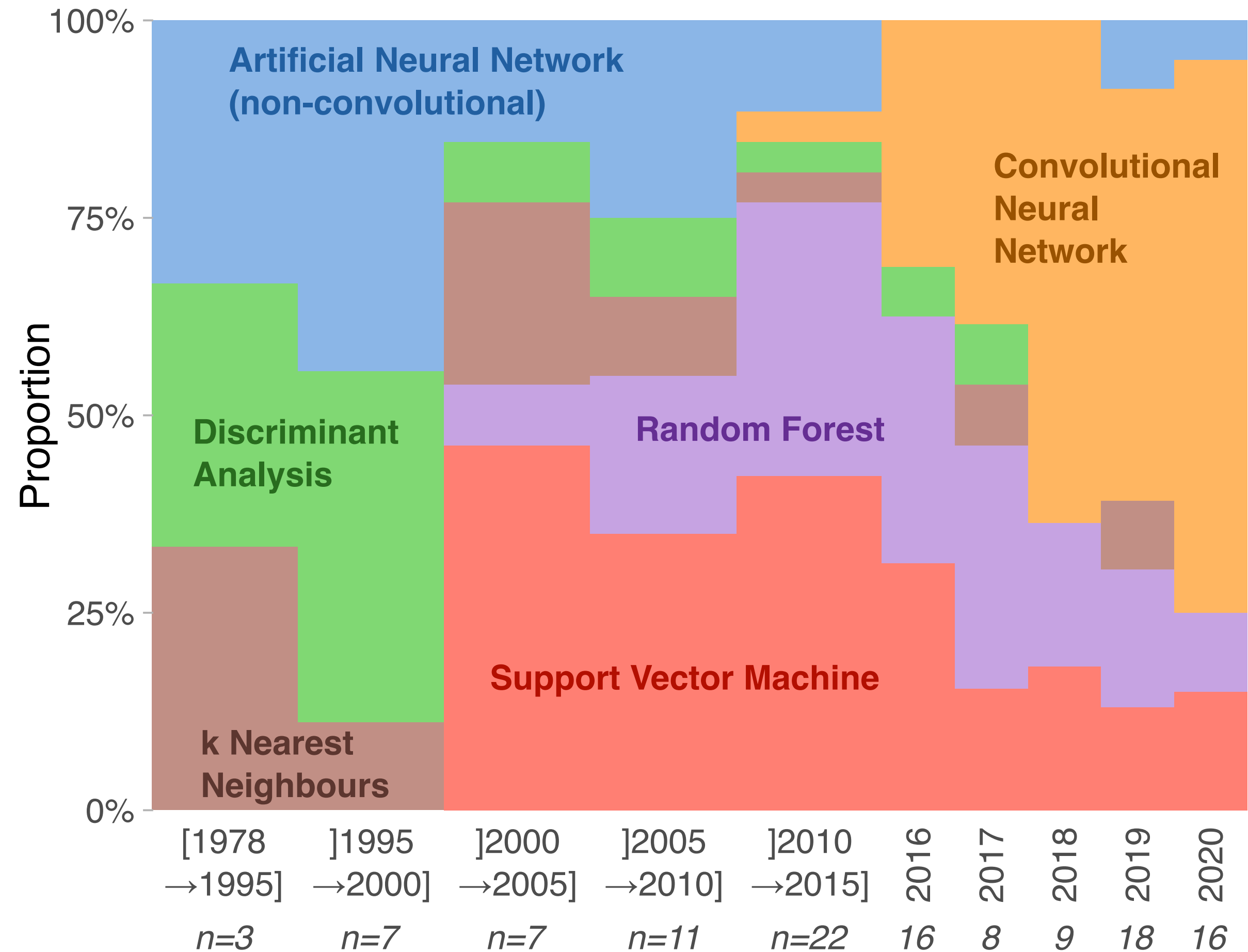
²¹Wageningen University and Research, Wageningen Marine Research, 1976 CP IJmuiden, The Netherlands



Machine learning to extract ecological information from observational data

Machine learning as an accelerator

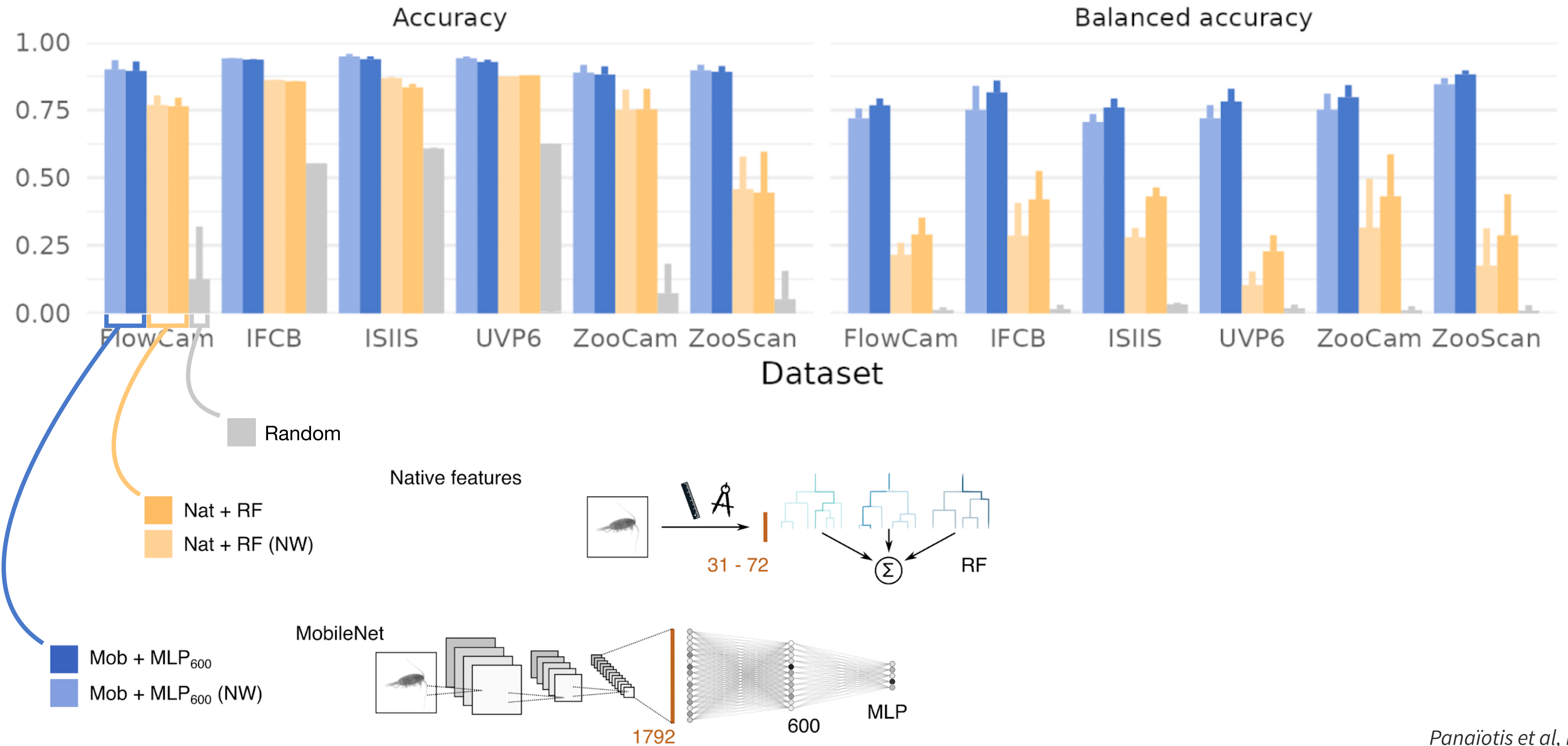
Historical approaches to (plankton) image classification



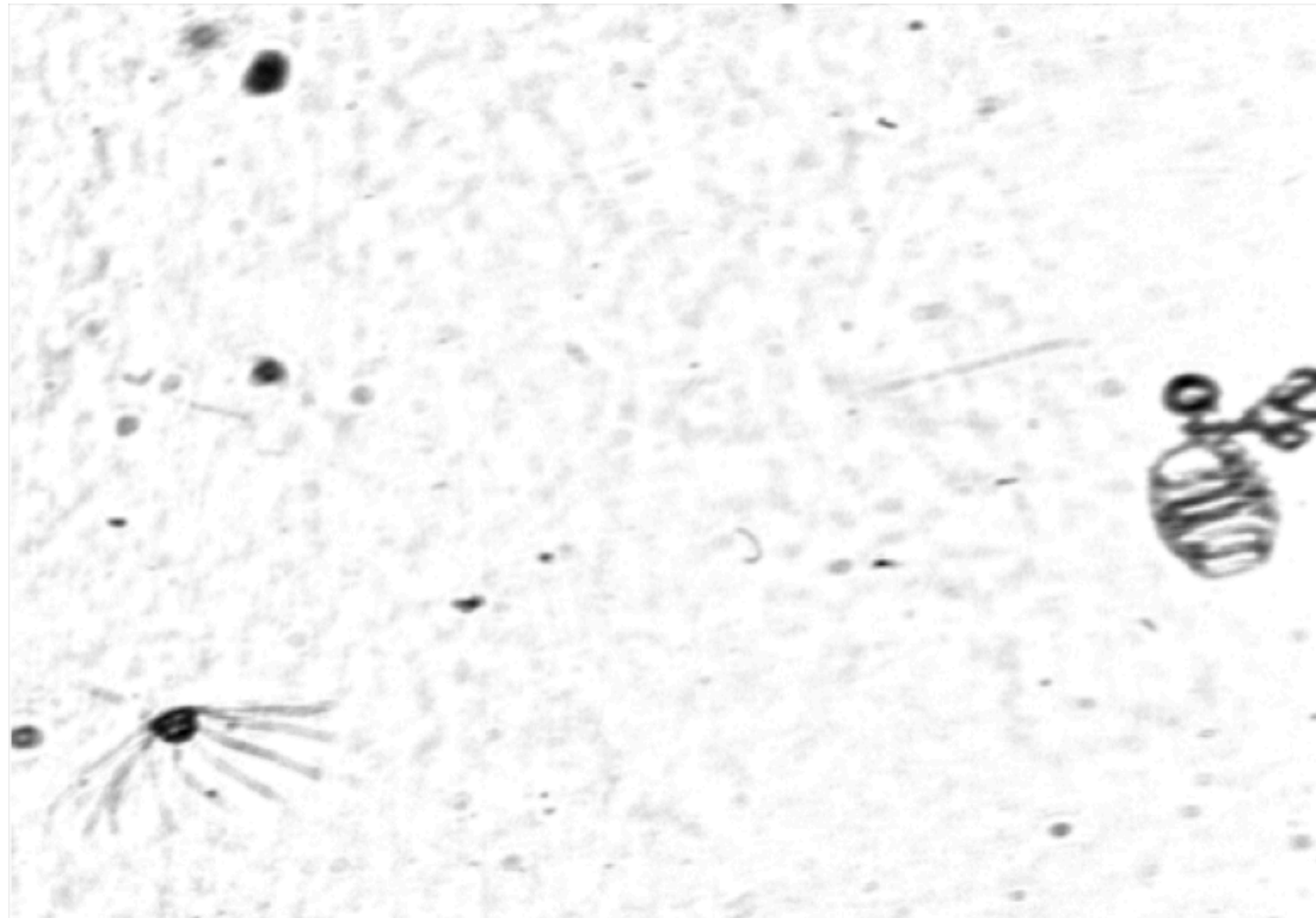
RandomForest was popular in technical papers but even more in actually implemented methods

Gradual **transition** from classical feature extraction + classification to CNNs

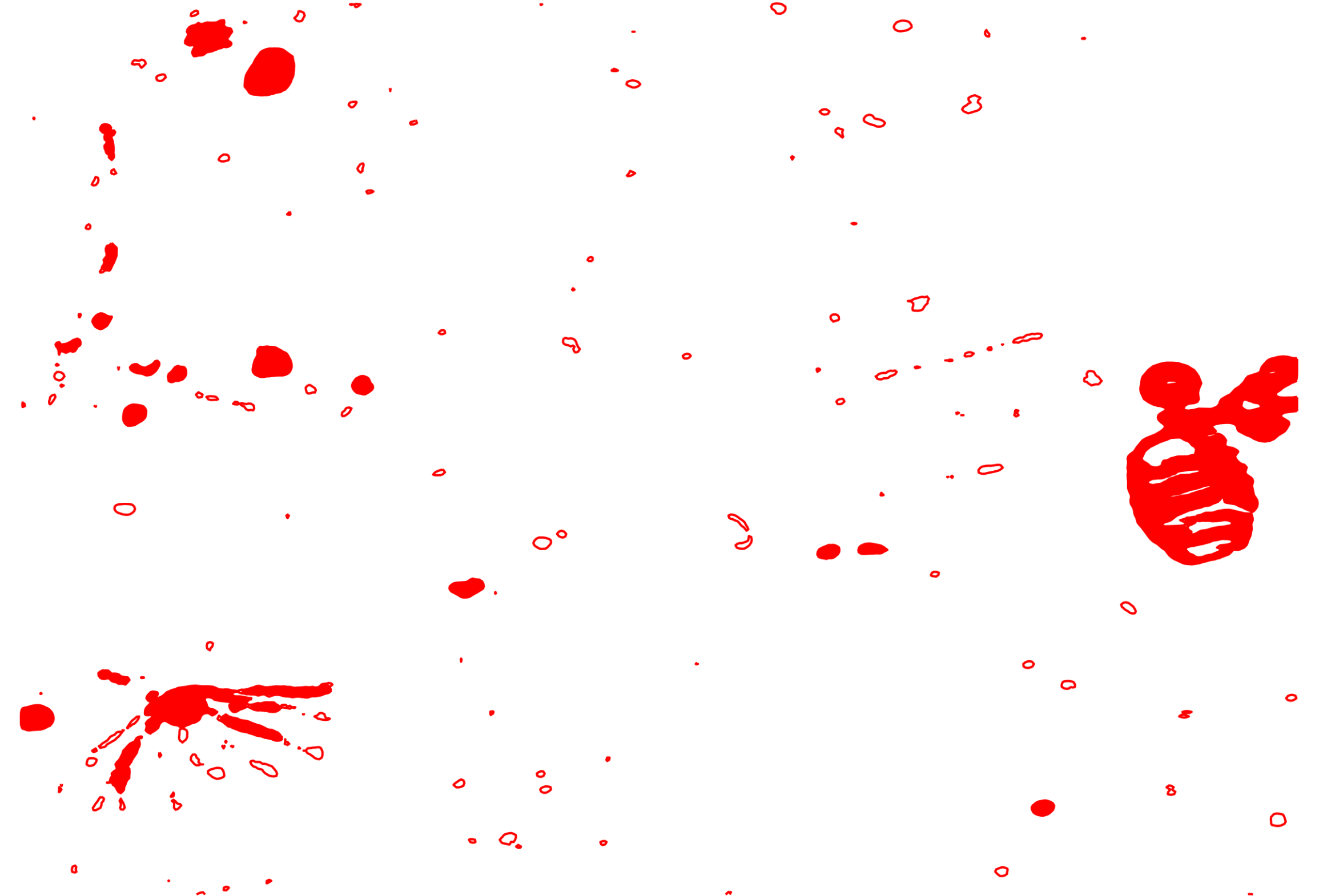
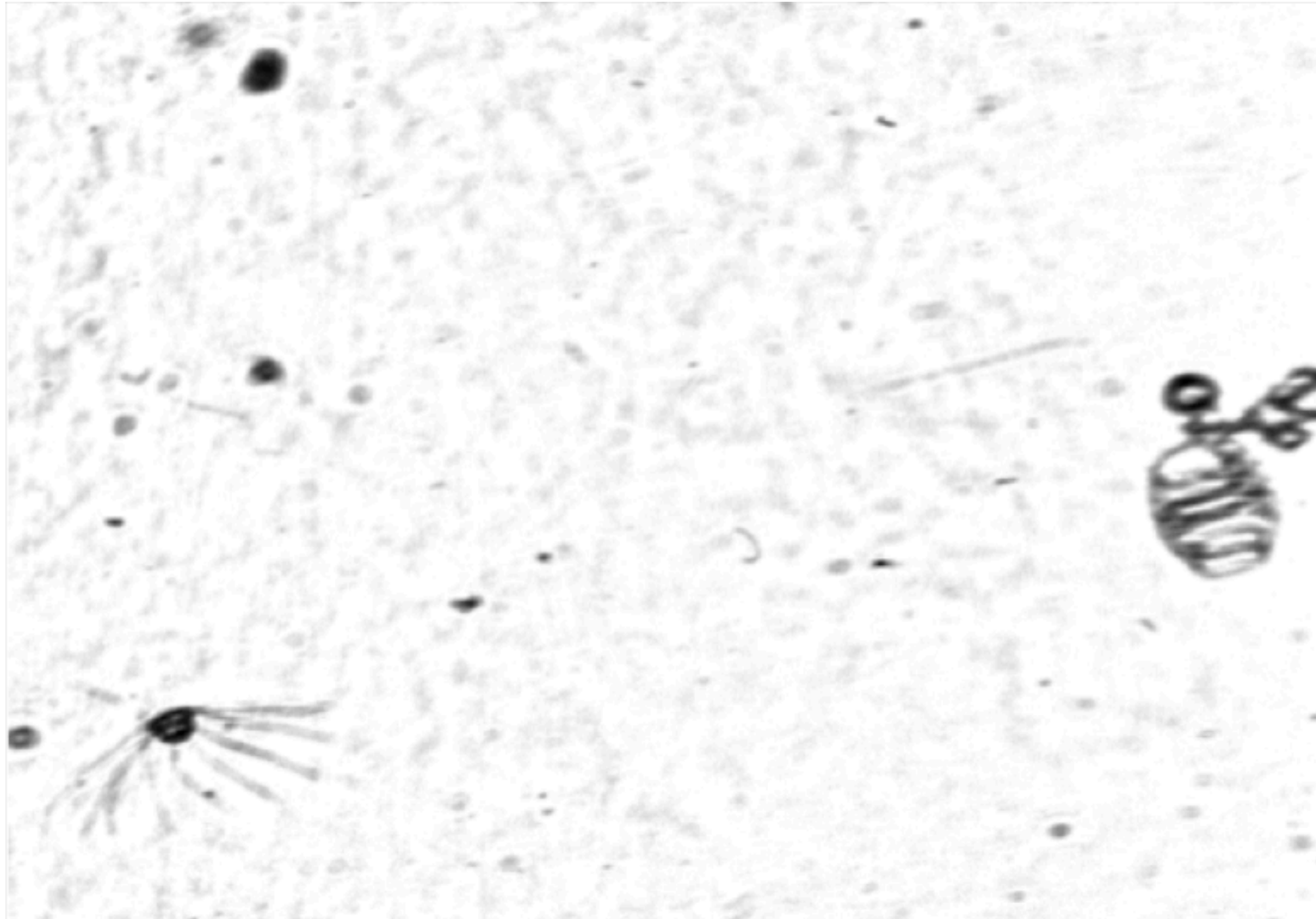
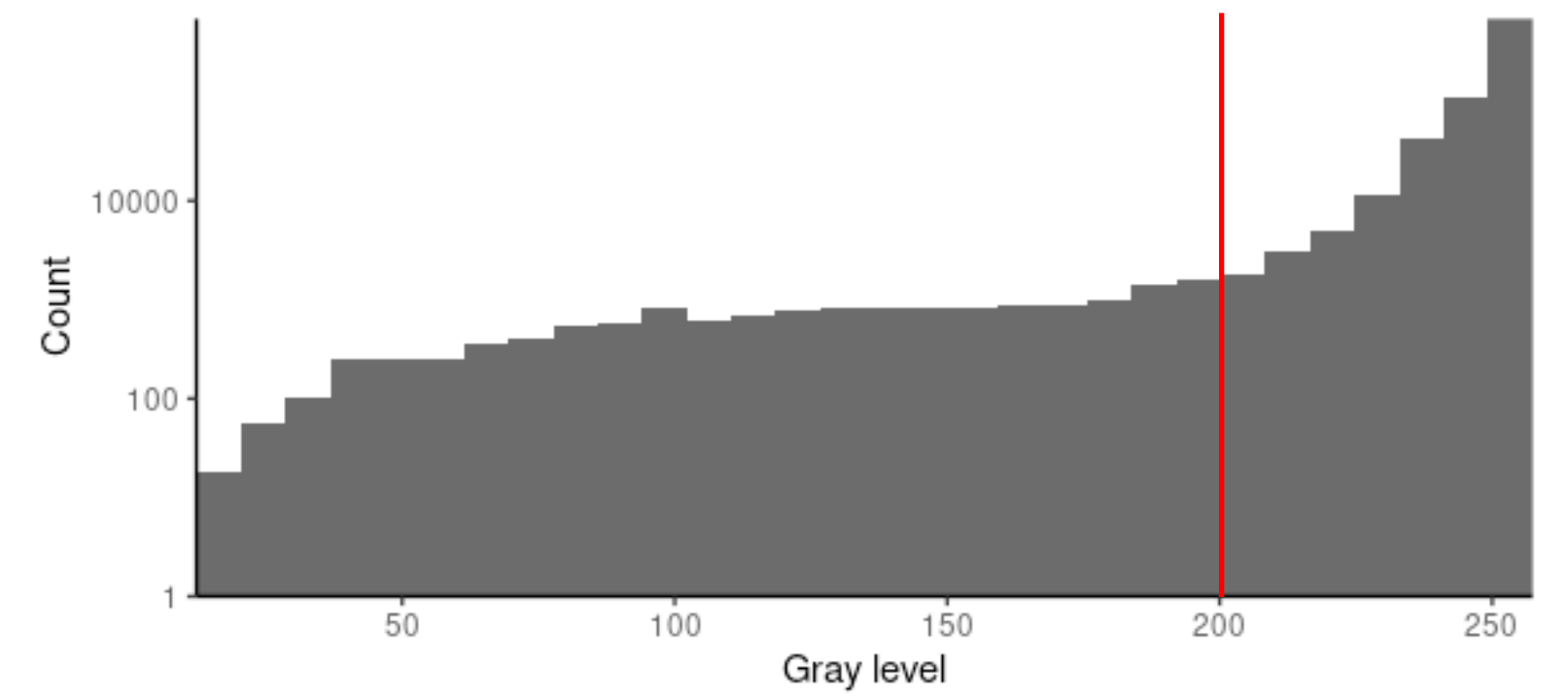
Classic vs deep: deep is better, for rare classes



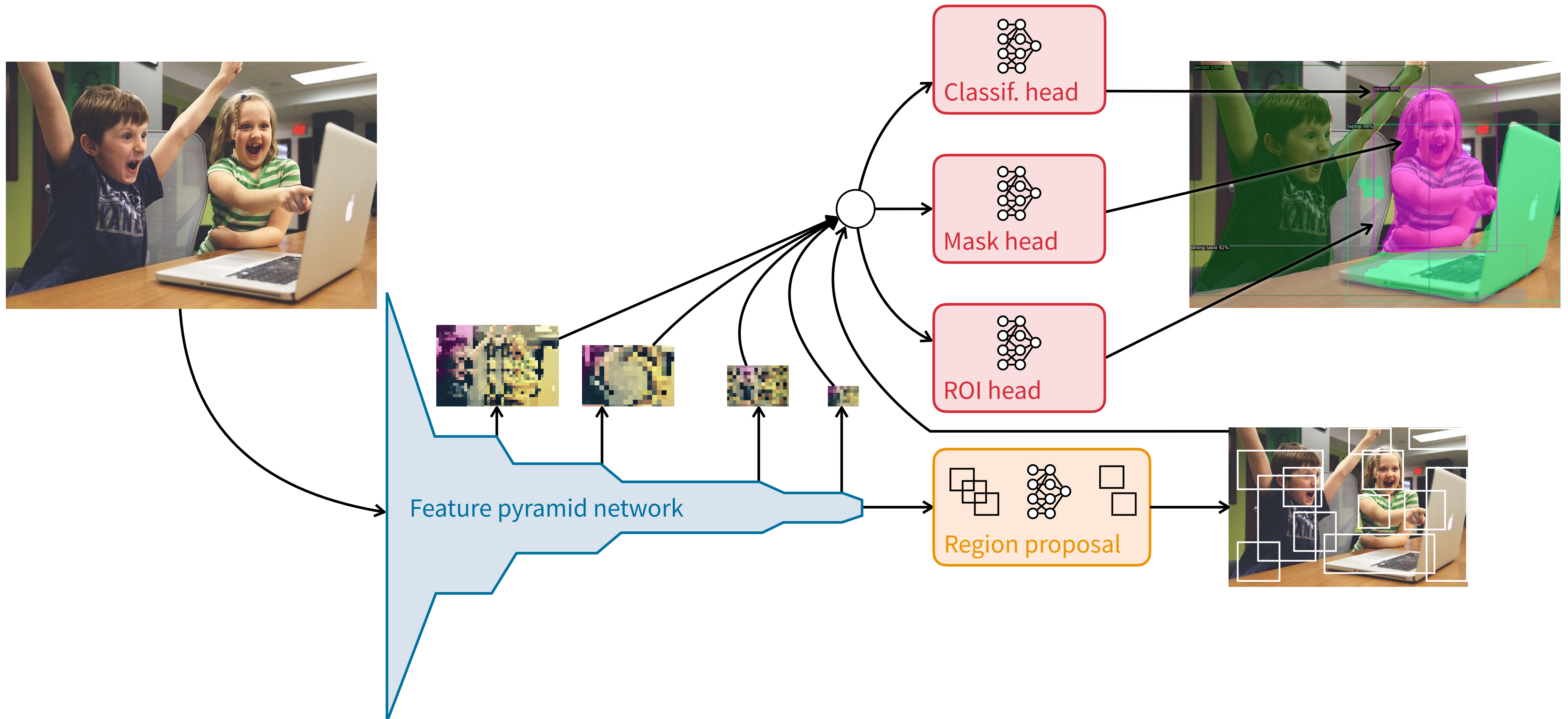
Classic segmentation: (adaptive) thresholding



Classic segmentation: (adaptive) thresholding

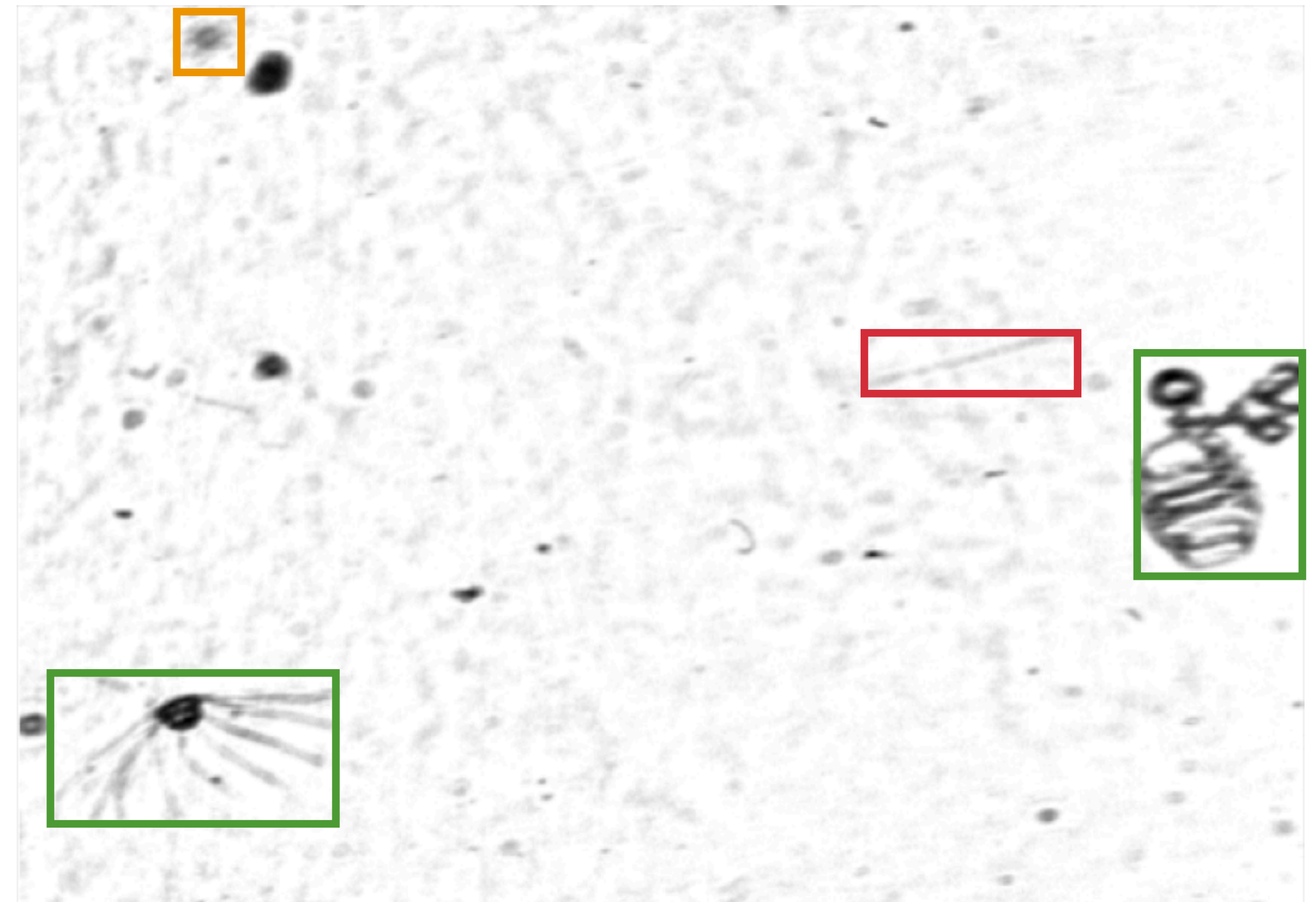
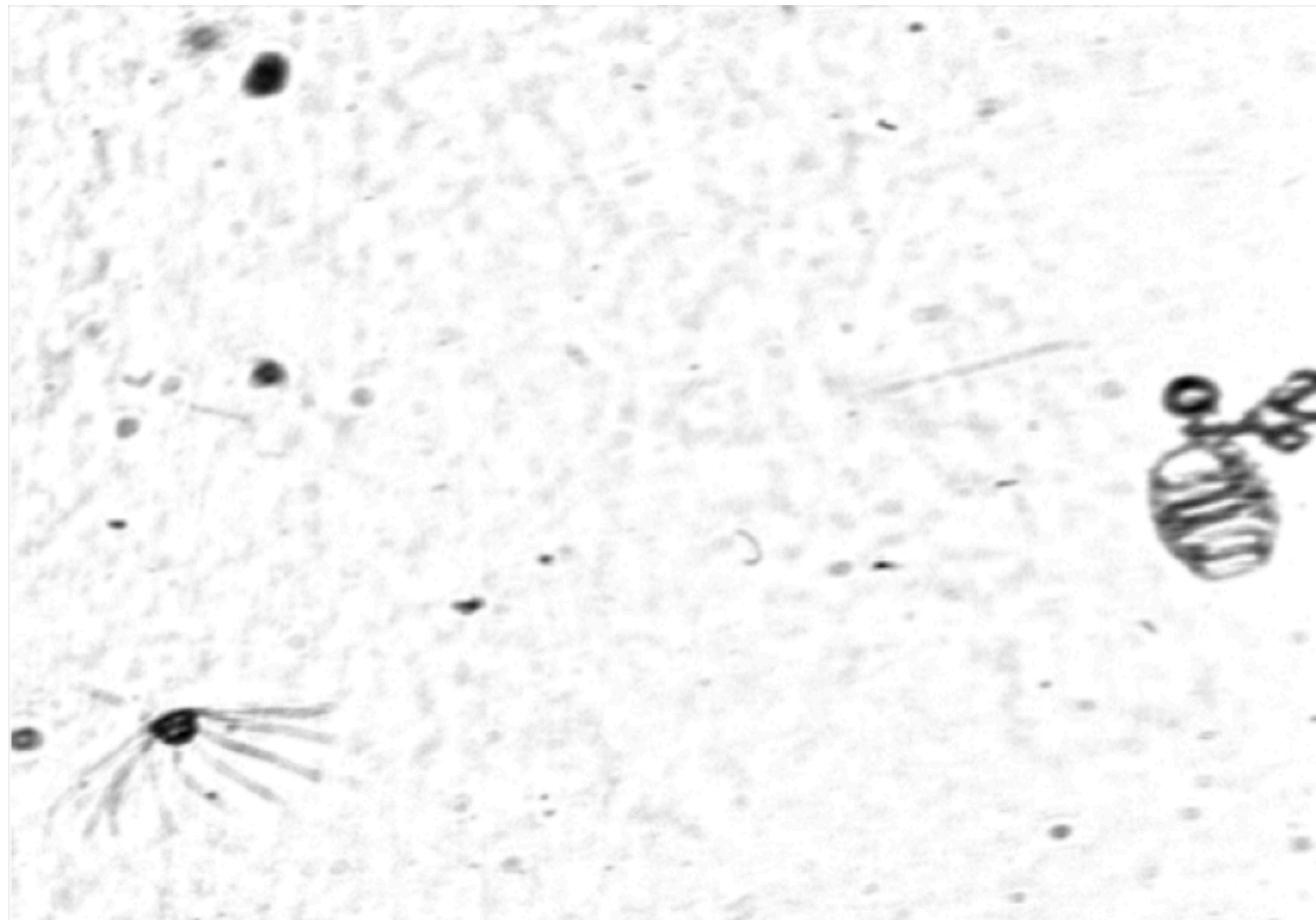


Deep object detection/segmentation:

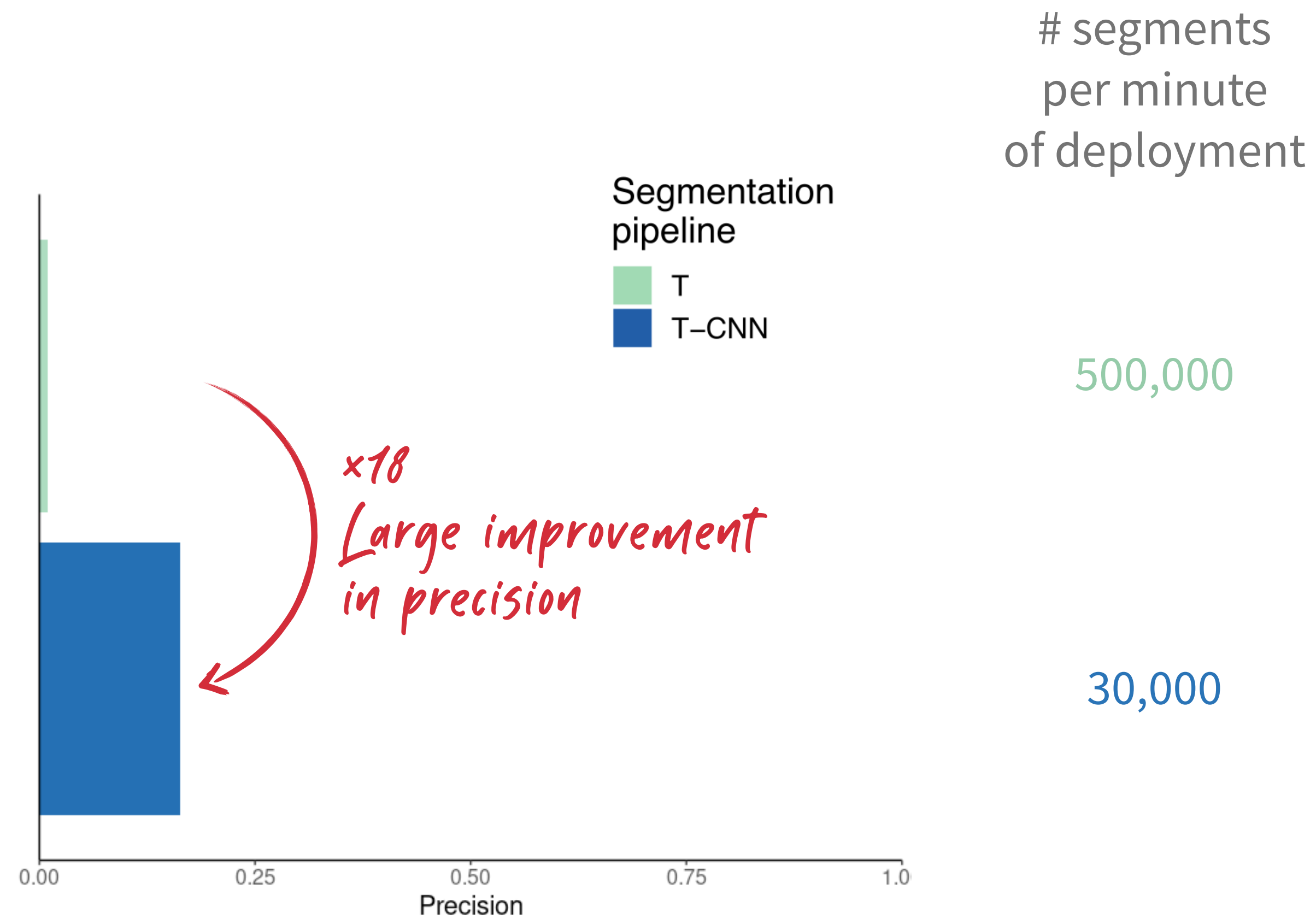


Deep object detection/segmentation:

 Detectron2

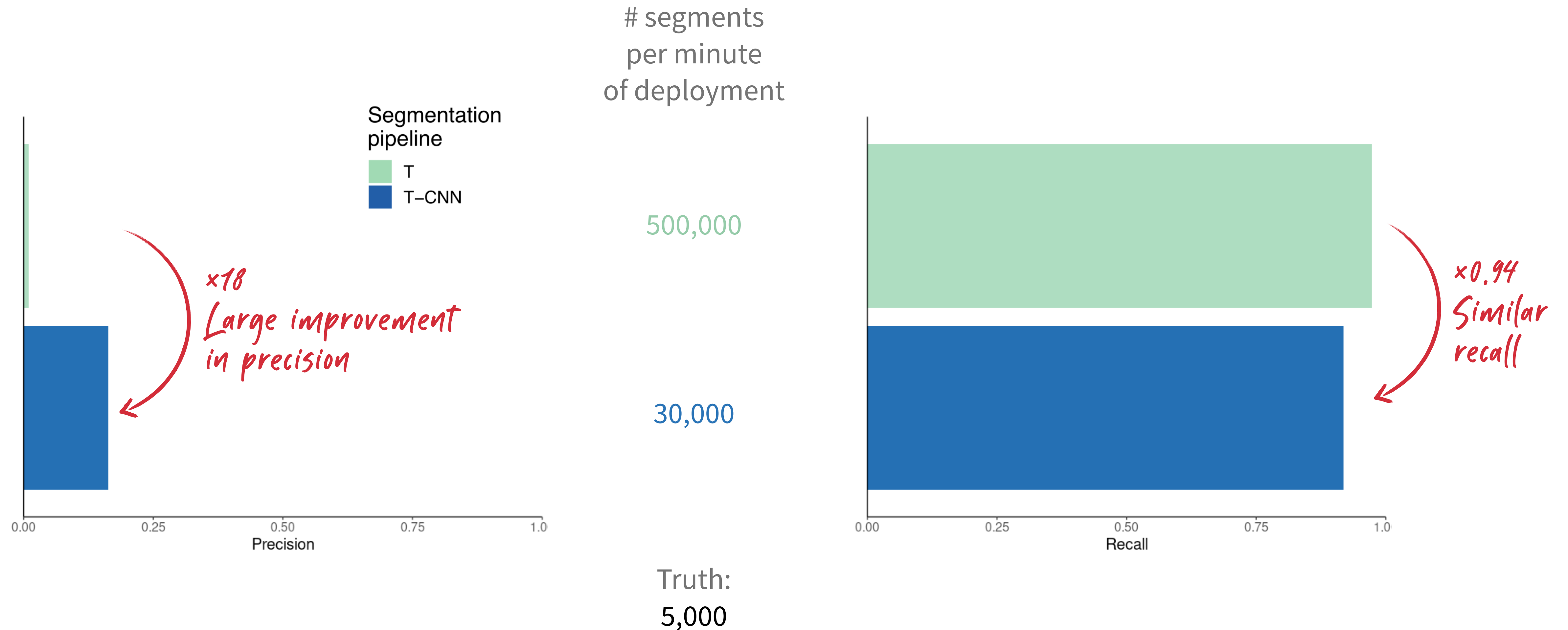


“Intelligent” segmentation uses context to improve precision



Truth:
5,000

“Intelligent” segmentation uses context to improve precision



Classification from passive acoustics: classic feature extraction

Feature extracted: distance between zero crossings

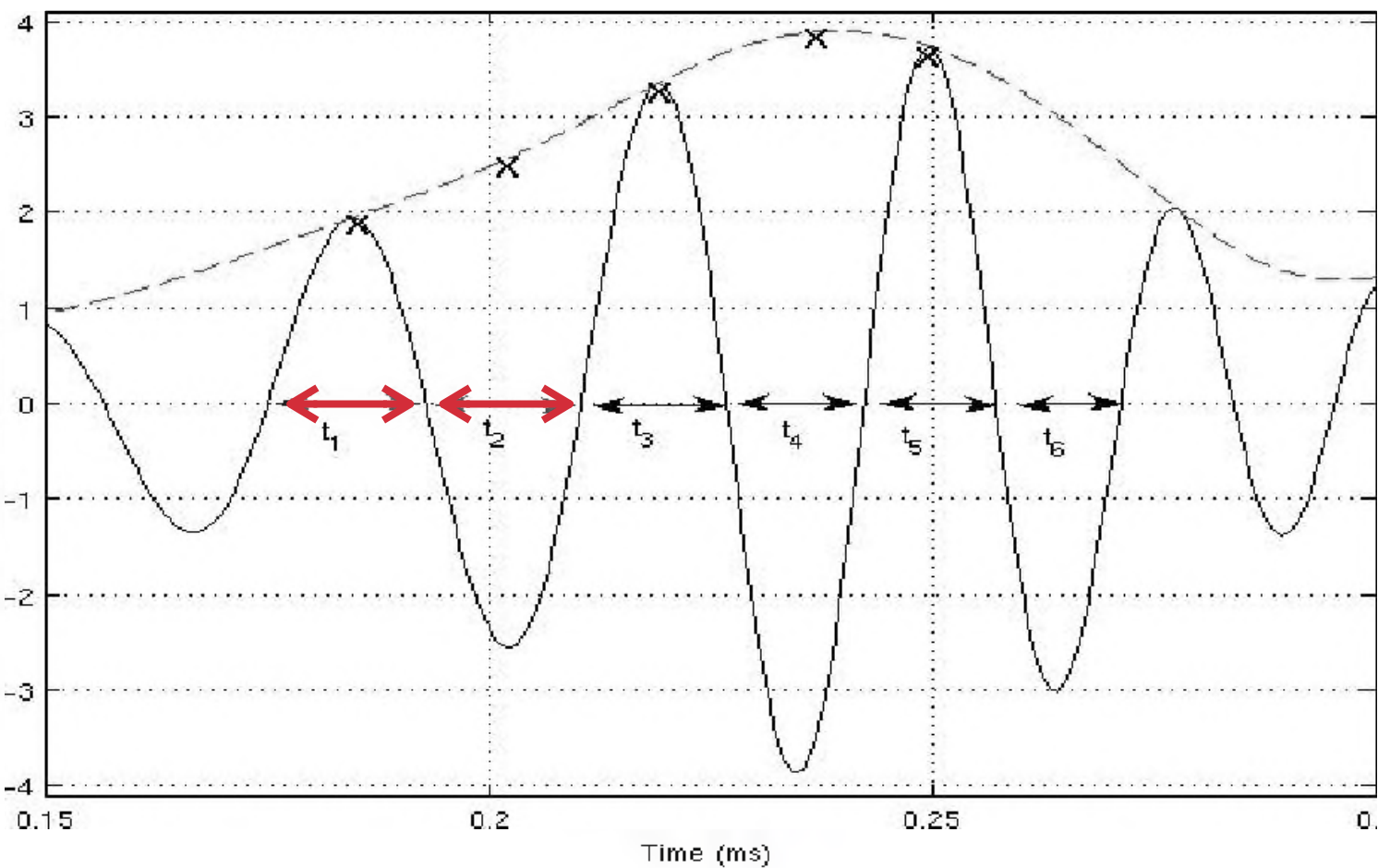


Figure 3: The times between zero-crossings and normalized envelope amplitudes were selected as features for classifying individual clicks.

Good separation for most signals

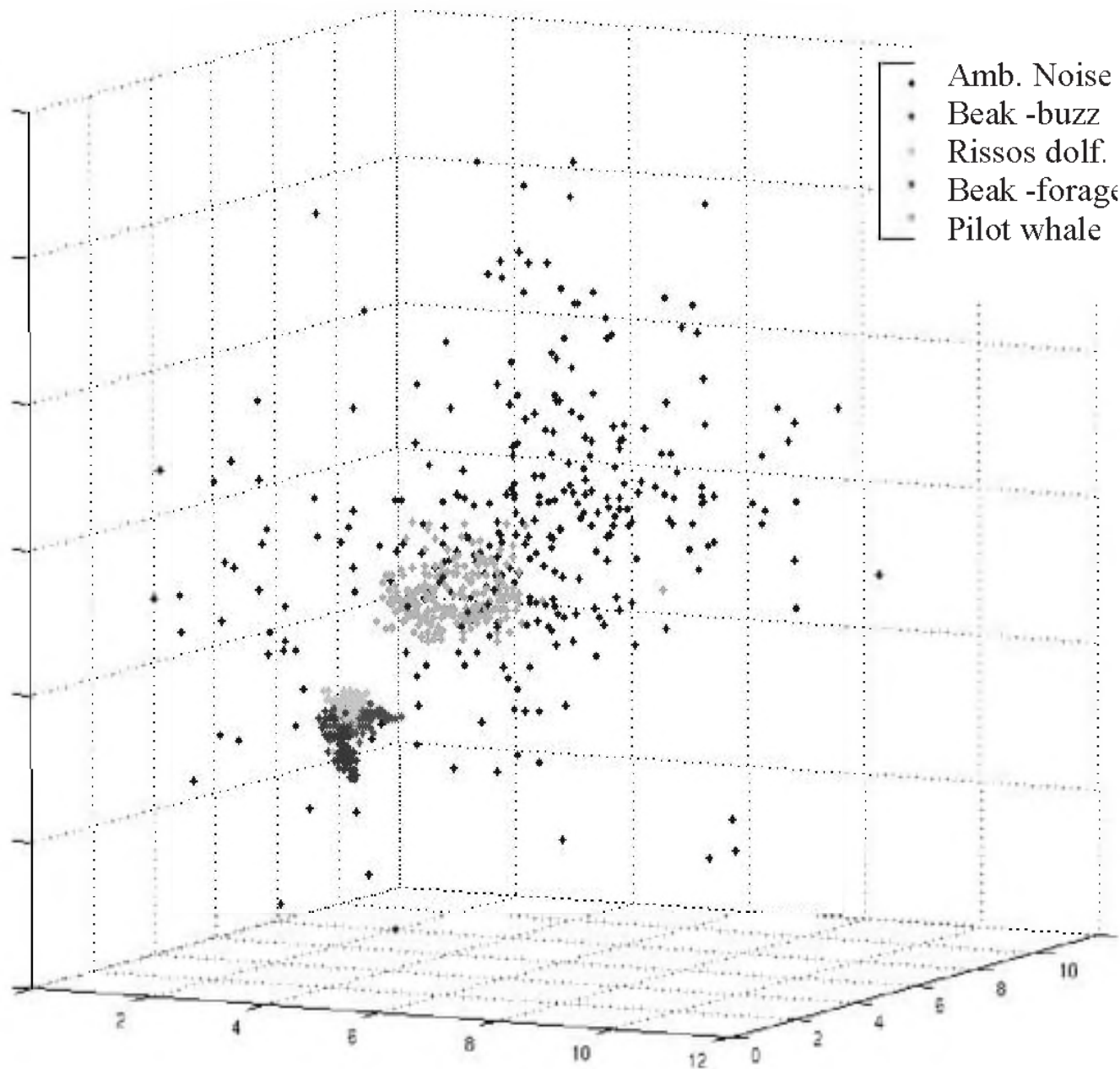


Figure 6: Distribution of the periods between three consecutive zero-crossings for ambient noise, *M. desirostris* (buzz), *Grampus griseus*, *M. desirostris* (forage), and *Globicephala macrorhynchus*.



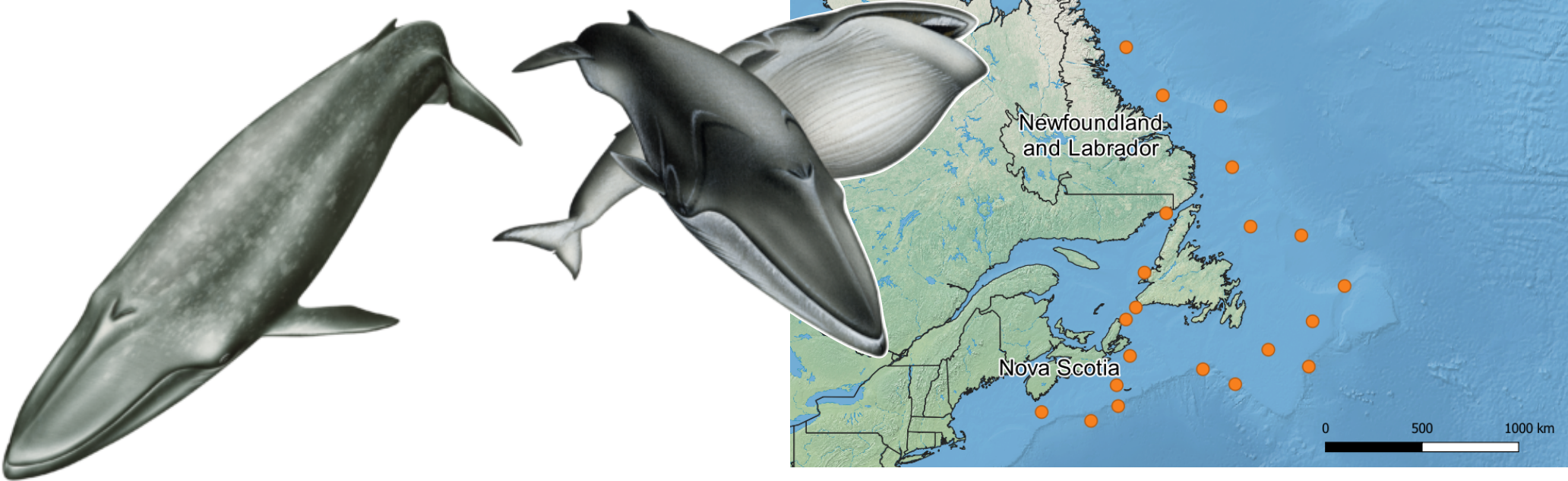
SVM for classification

Good performance on training set

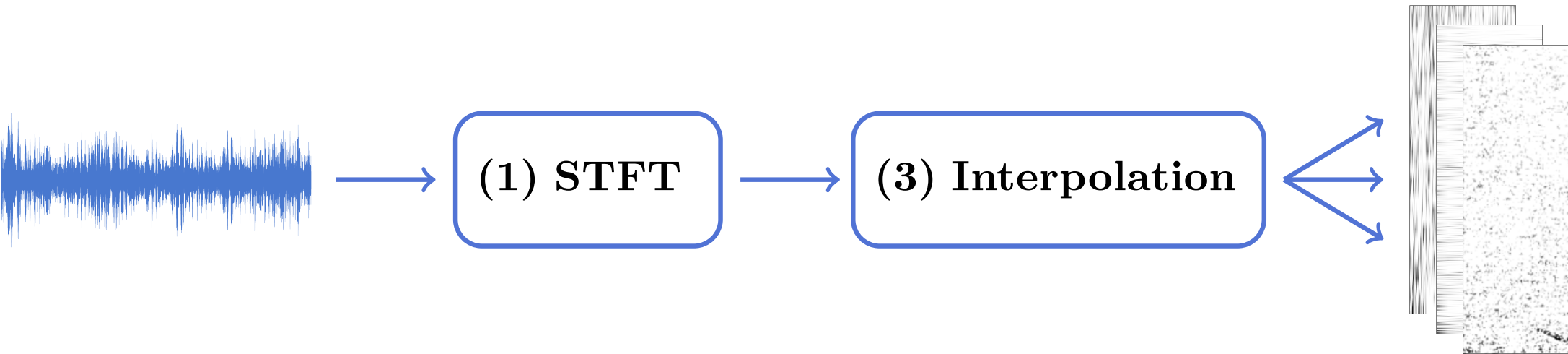
Class	P_{cc}	P_{miss}	P_{nse}
<i>Mesoplodon</i> (forage)	1.0000	0.0000	0.9680
<i>Mesoplodon</i> (buzz)	0.7900	0.2100	0.9600
<i>Globicephala</i>	0.9380	0.0620	0.9682
<i>Grampus</i>	0.9451	0.0549	0.9721

Does **not generalise** well (in particular to species not seen during training)

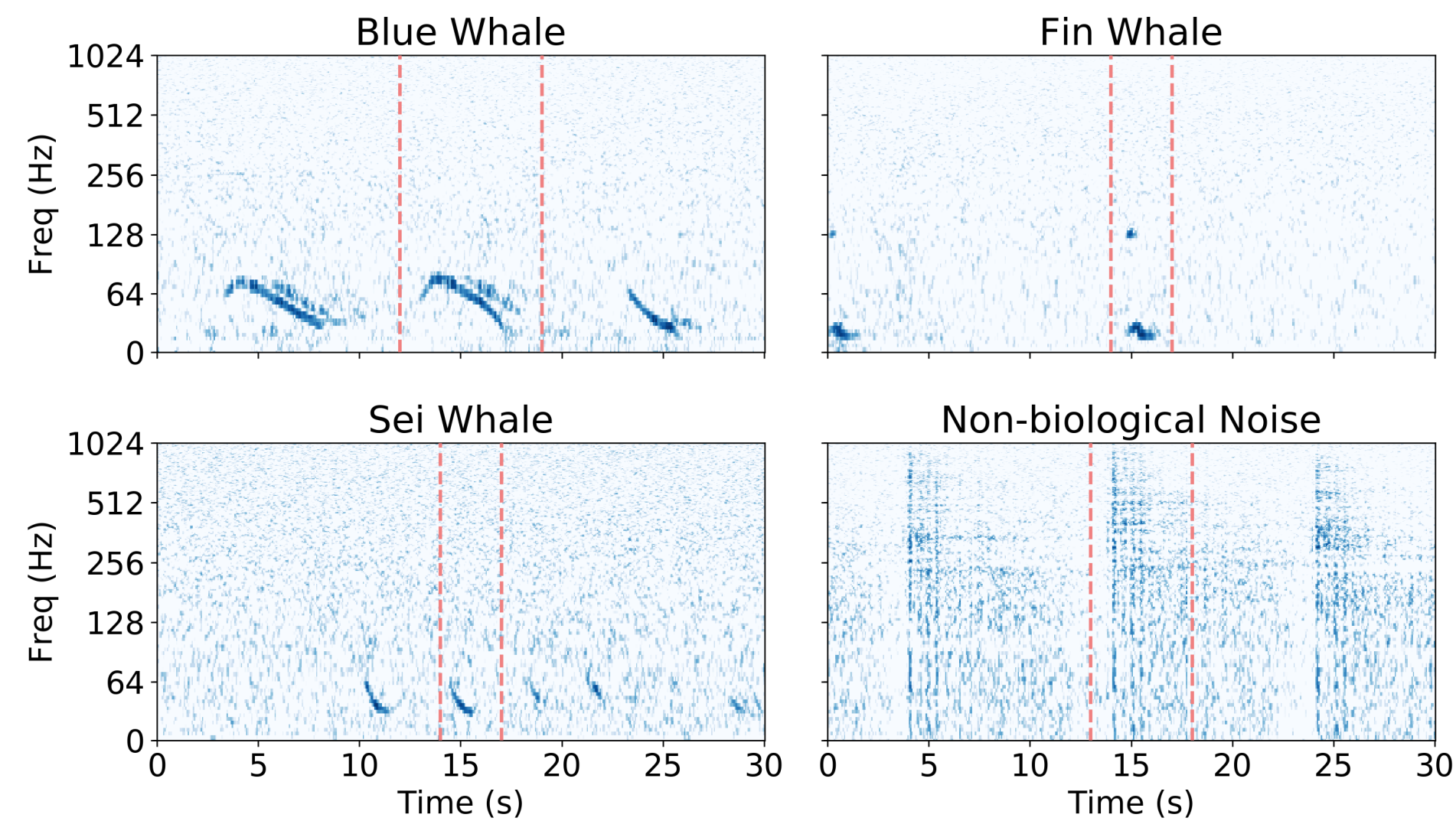
Classification from passive acoustics: deep approach



Sound converted to **spectrogram** using Short Term Fourier Transform with various parameters



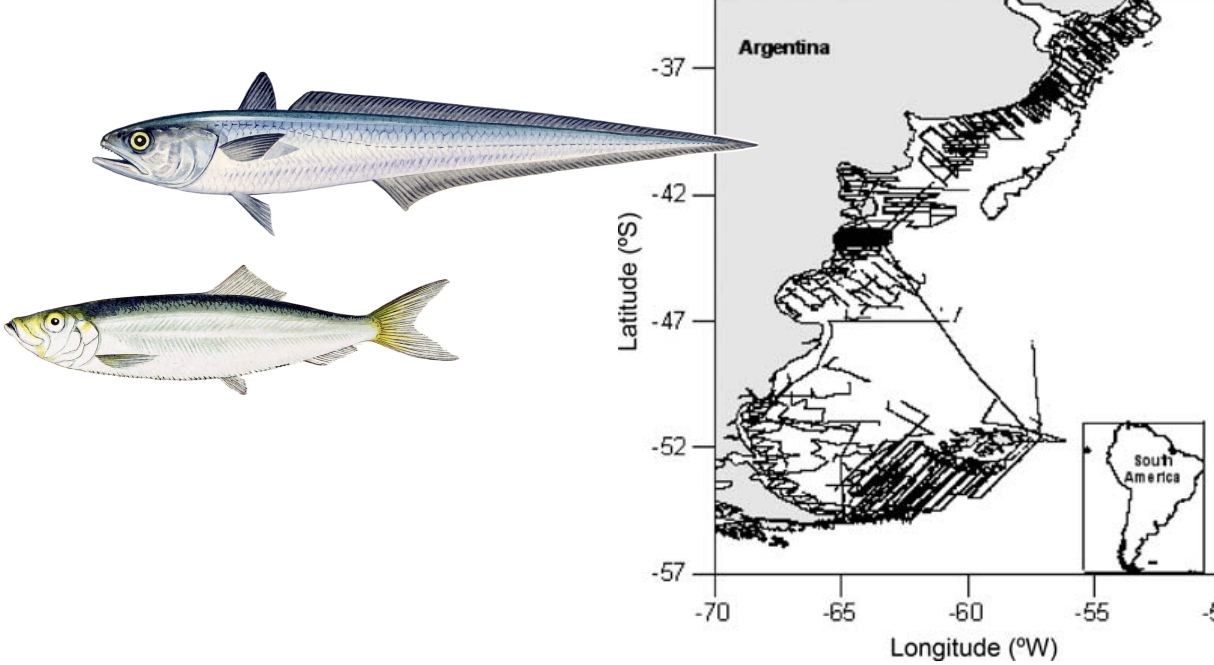
Detection and annotation by **humans**



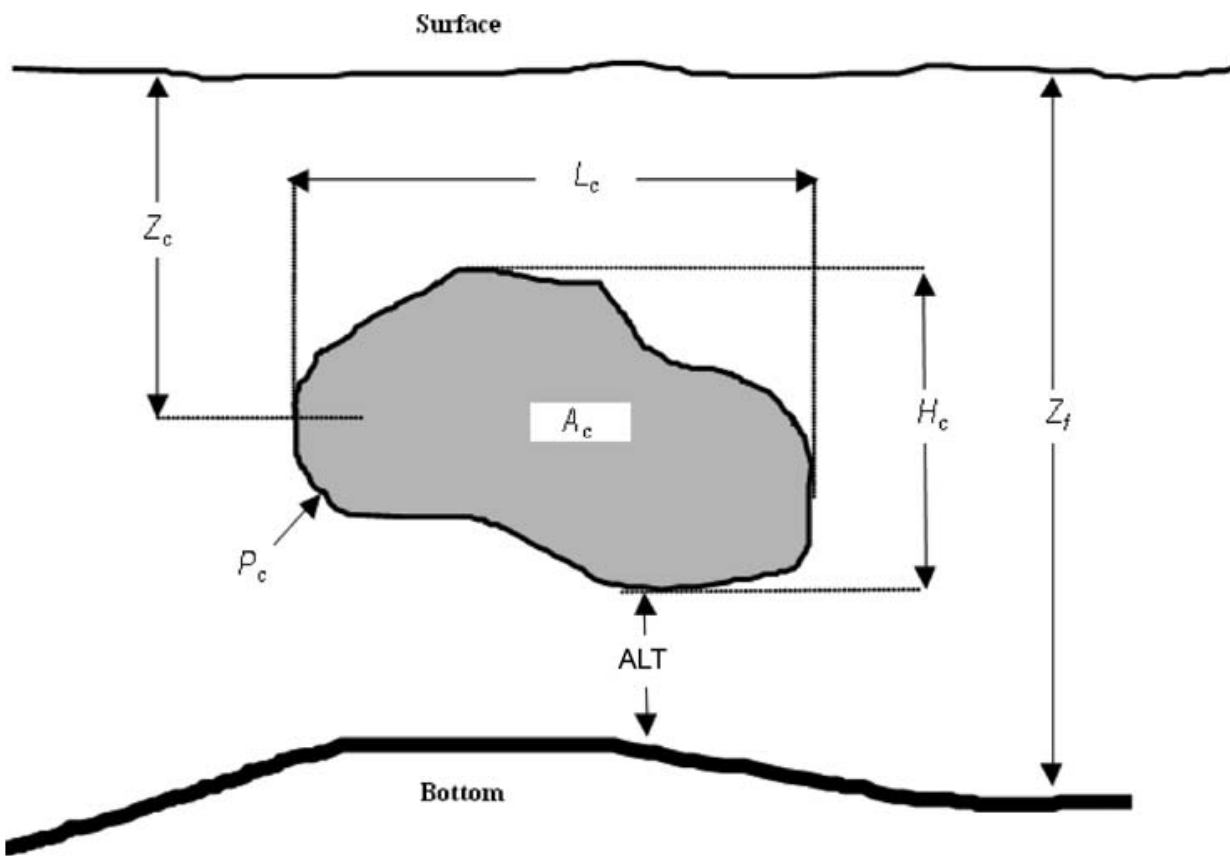
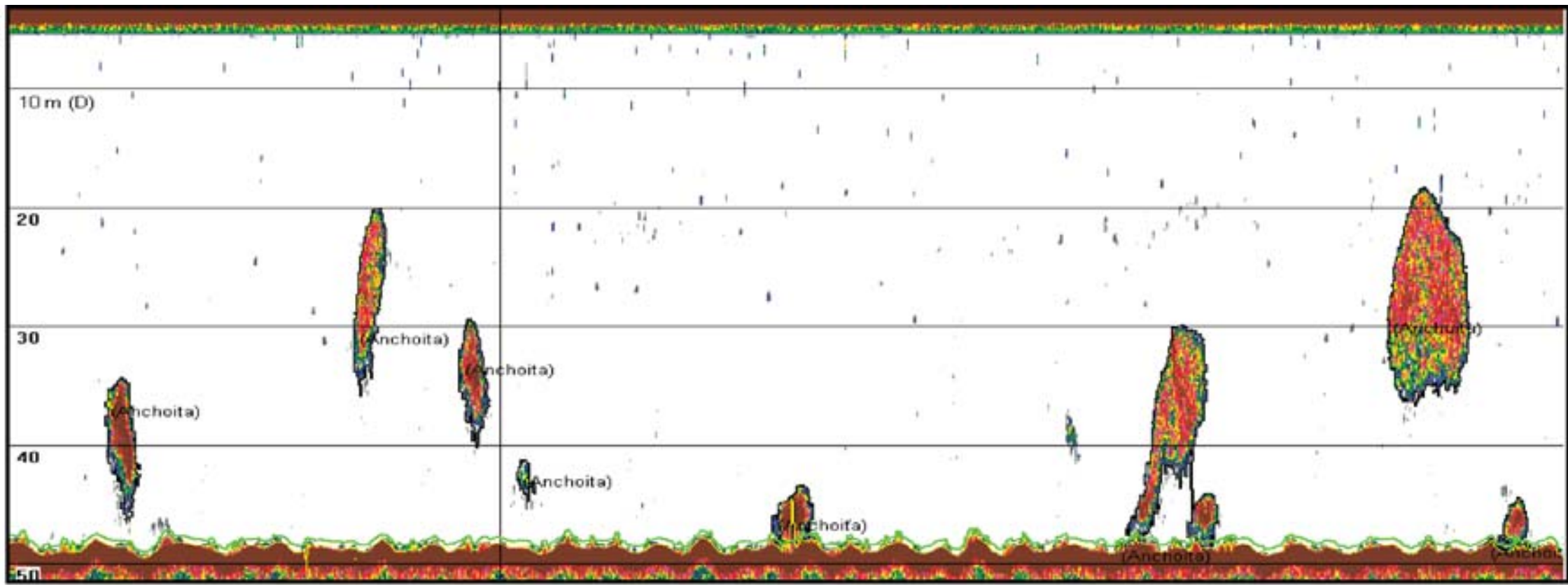
Test of classification based on human annotation, from the **image of the spectrogram**, with two CNNs

ResNet-50 Performance					
	NFFT	Accuracy	Precision	Recall	F-1 Score
3-channels (Hz)	-	0.953 (± 0.016)	0.887 (± 0.045)	0.871 (± 0.036)	0.878 (± 0.031)
	256	0.883 (± 0.022)	0.714 (± 0.060)	0.641 (± 0.037)	0.675 (± 0.046)
1-channel (Hz)	2048	0.944 (± 0.009)	0.863 (± 0.036)	0.838 (± 0.039)	0.850 (± 0.023)
	16384	0.943 (± 0.013)	0.860 (± 0.032)	0.847 (± 0.058)	0.853 (± 0.031)
1-channel (mels)	2048	0.895 (± 0.031)	0.762 (± 0.067)	0.723 (± 0.048)	0.742 (± 0.044)
VGG-19 Performance					
	NFFT	Accuracy	Precision	Recall	F-1 Score
3-channels (Hz)	-	0.961 (± 0.017)	0.906 (± 0.044)	0.892 (± 0.049)	0.899 (± 0.041)
	256	0.914 (± 0.024)	0.790 (± 0.048)	0.771 (± 0.070)	0.780 (± 0.053)
1-channel (Hz)	2048	0.959 (± 0.019)	0.899 (± 0.041)	0.889 (± 0.048)	0.894 (± 0.039)
	16384	0.951 (± 0.017)	0.871 (± 0.037)	0.878 (± 0.038)	0.875 (± 0.028)
1-channel (mels)	2048	0.918 (± 0.022)	0.818 (± 0.043)	0.784 (± 0.036)	0.801 (± 0.034)

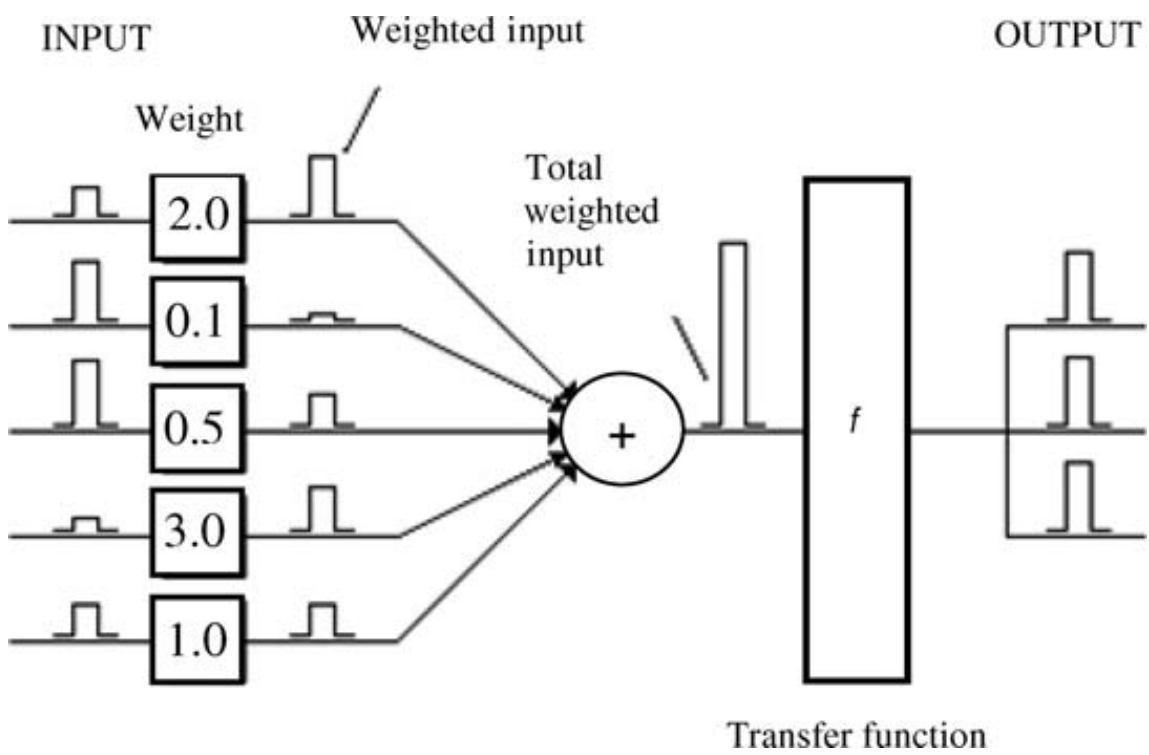
Target classification from active acoustics: classic approach



Extraction of various **features** of each school



Classification using an Artificial Neural Network

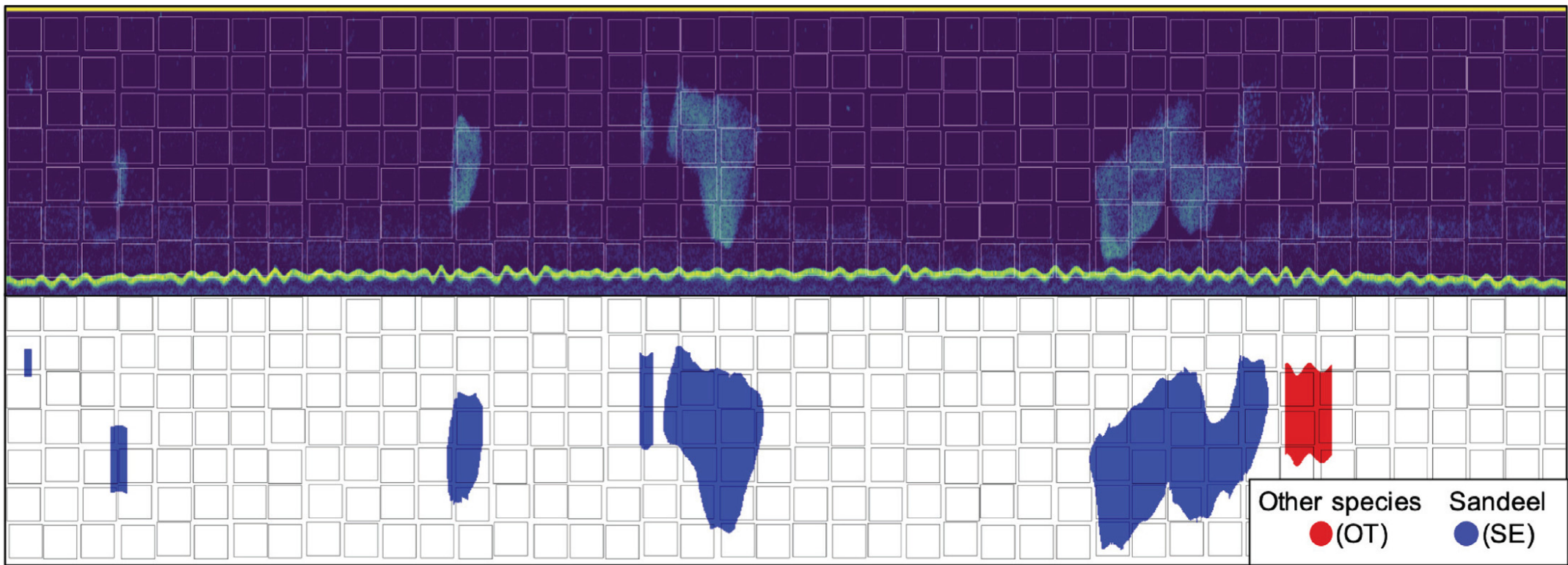


		Recognized category (%)					
FF		Bonaerensis anchovy	Sprat	Rough scad	Hoki	Blue whiting	Patagonian anchovy
True category (%)	Bonaerensis anchovy	93.3	0	0.7	0.9	0.1	5.1
	Sprat	41.8	35.4	0	6.3	1.3	15.2
	Rough scad	63.5	0	36.5	0	0	0
	Hoki	2.3	0.2	0.2	92.7	0.8	3.9
	Blue whiting	0	2.6	0	23.7	73.7	0
	Patagonian anchovy	27.1	1.6	0	2.3	0	69.0

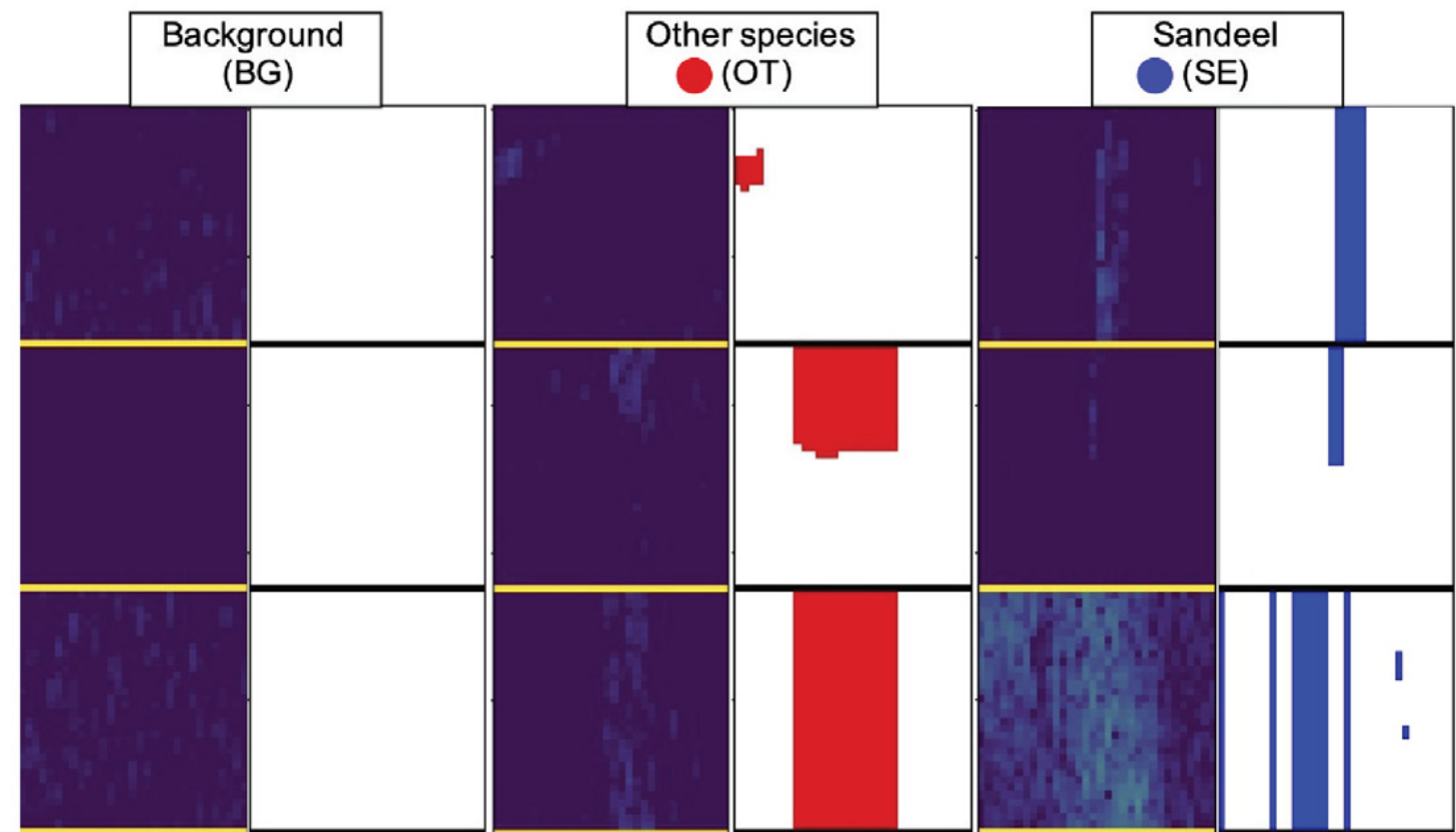
Target classification from active acoustics: semi-supervised deep learning



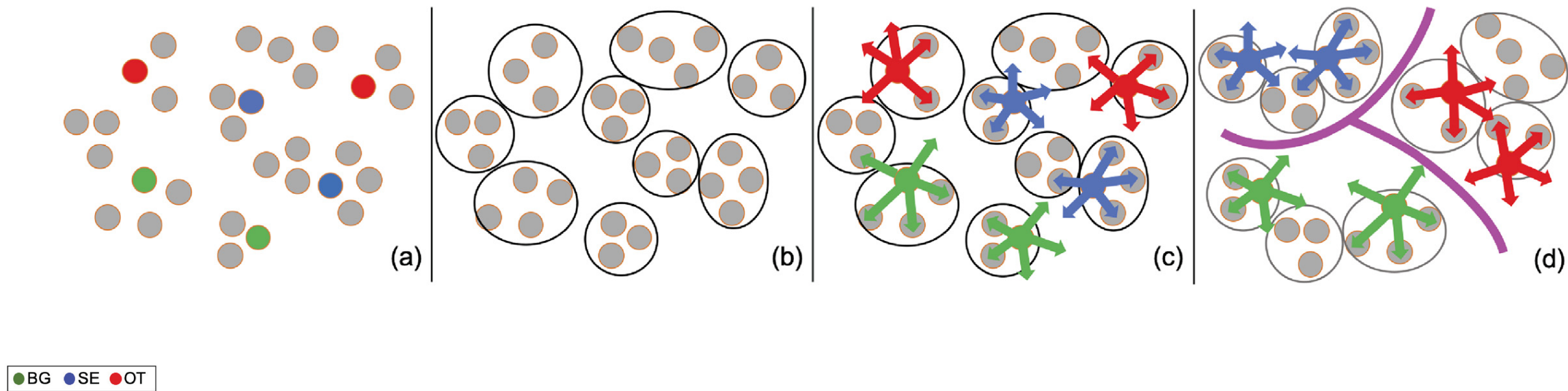
Cut input in 32×32×4 **blocs**



Annotate some blocs



Cluster blocs based on similarity, **propagate** annotations to similar blocs



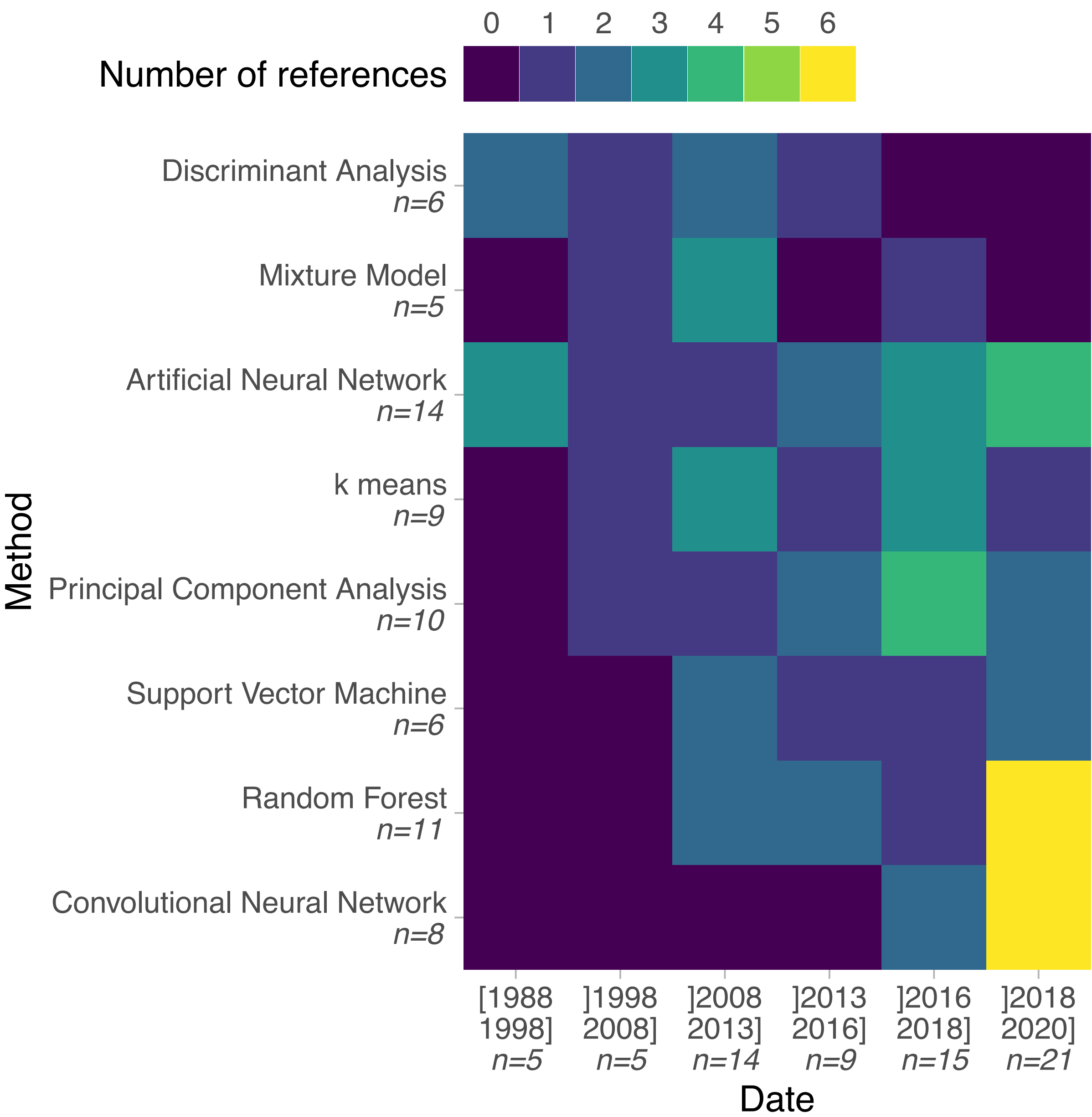
Train CNN on propagated labels: **good** performance even with very **few annotations**

Class-bal. test set Annotation ratio	Accuracy					
	Semi-supervised				Supervised	
	U-Im Ours	S3VM	U-Ba Ours	S3VM	AO Ours	SDL
1.000					0.8202	0.8000
0.100	0.7814	0.7341	0.7896	0.7340	0.7531	0.7496
0.050	0.7484	0.6668	0.7694	0.6668	0.6899	0.6909
0.025	0.7364	0.5838	0.7159	0.5827	0.6495	0.6108

Transition towards deep learning for acoustic target classification

Gradual in time

Like **most** other fields!

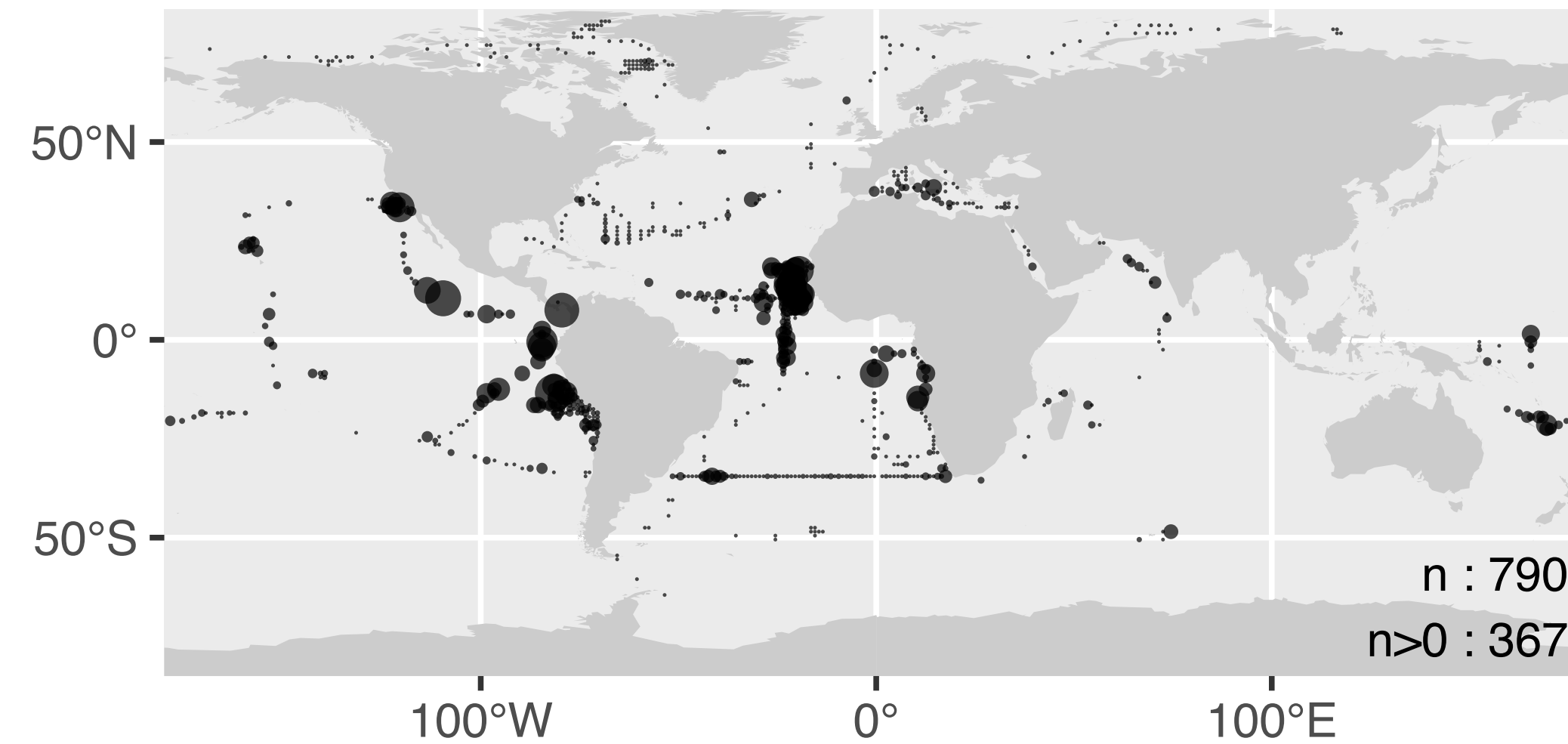


The background of the slide features a stylized, high-contrast image of a tree. The tree's branches and leaves are rendered in a light blue and white color scheme, creating a delicate, almost ethereal appearance. The tree is positioned on the right side of the frame, with its branches extending towards the left. The overall aesthetic is clean and modern, with a focus on natural elements.

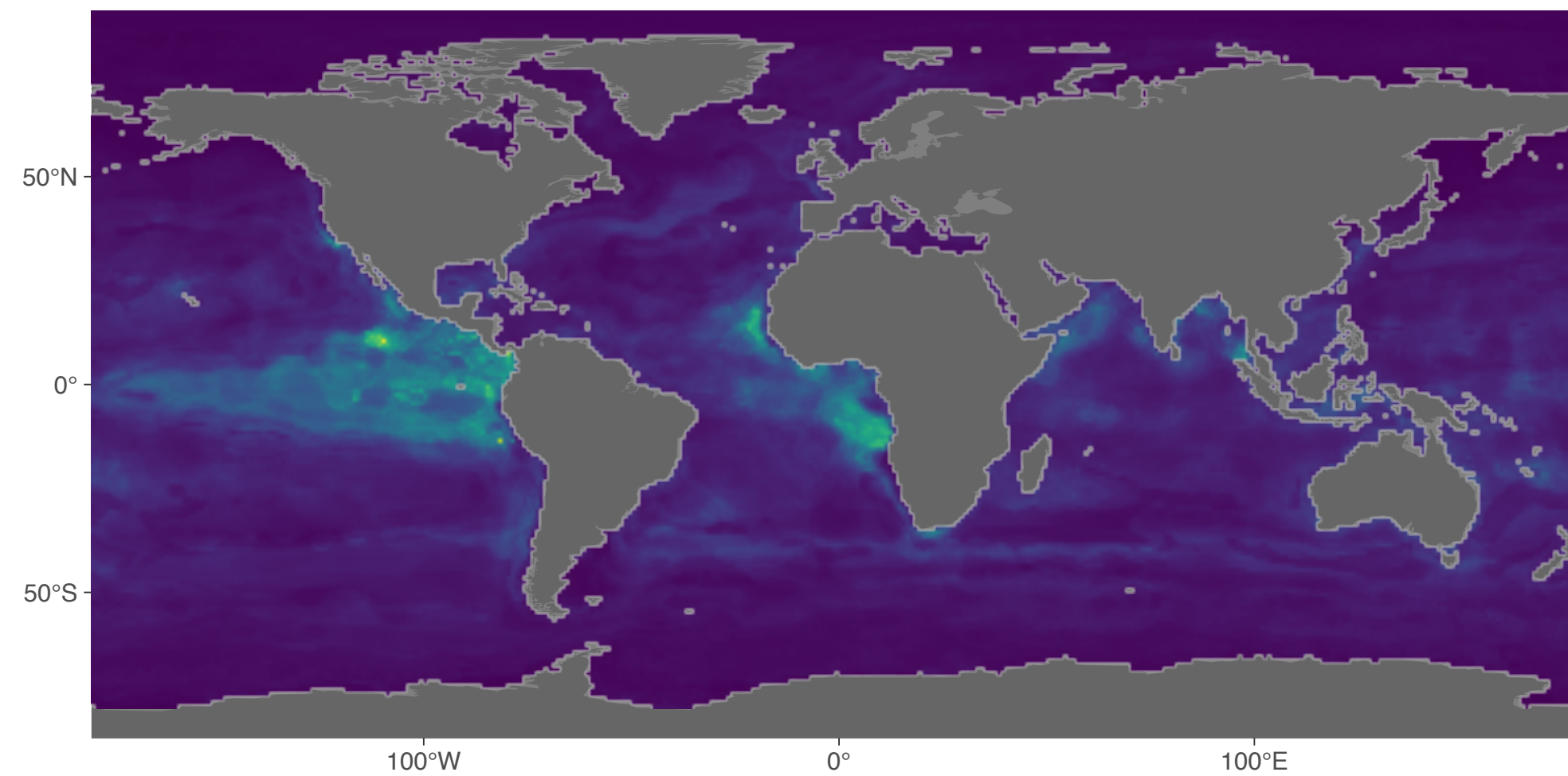
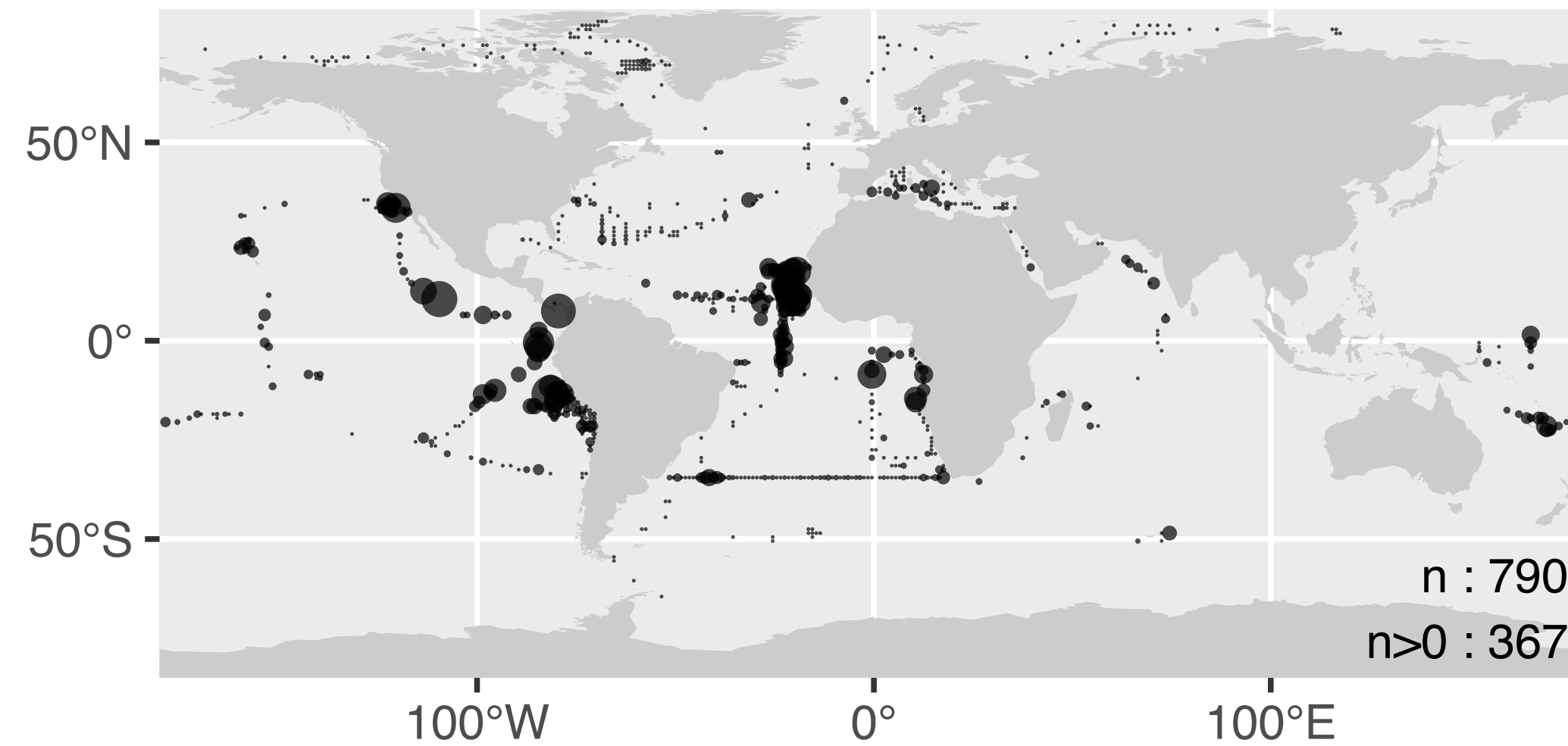
Machine learning to improve ecological understanding

Machine learning as an investigative tool

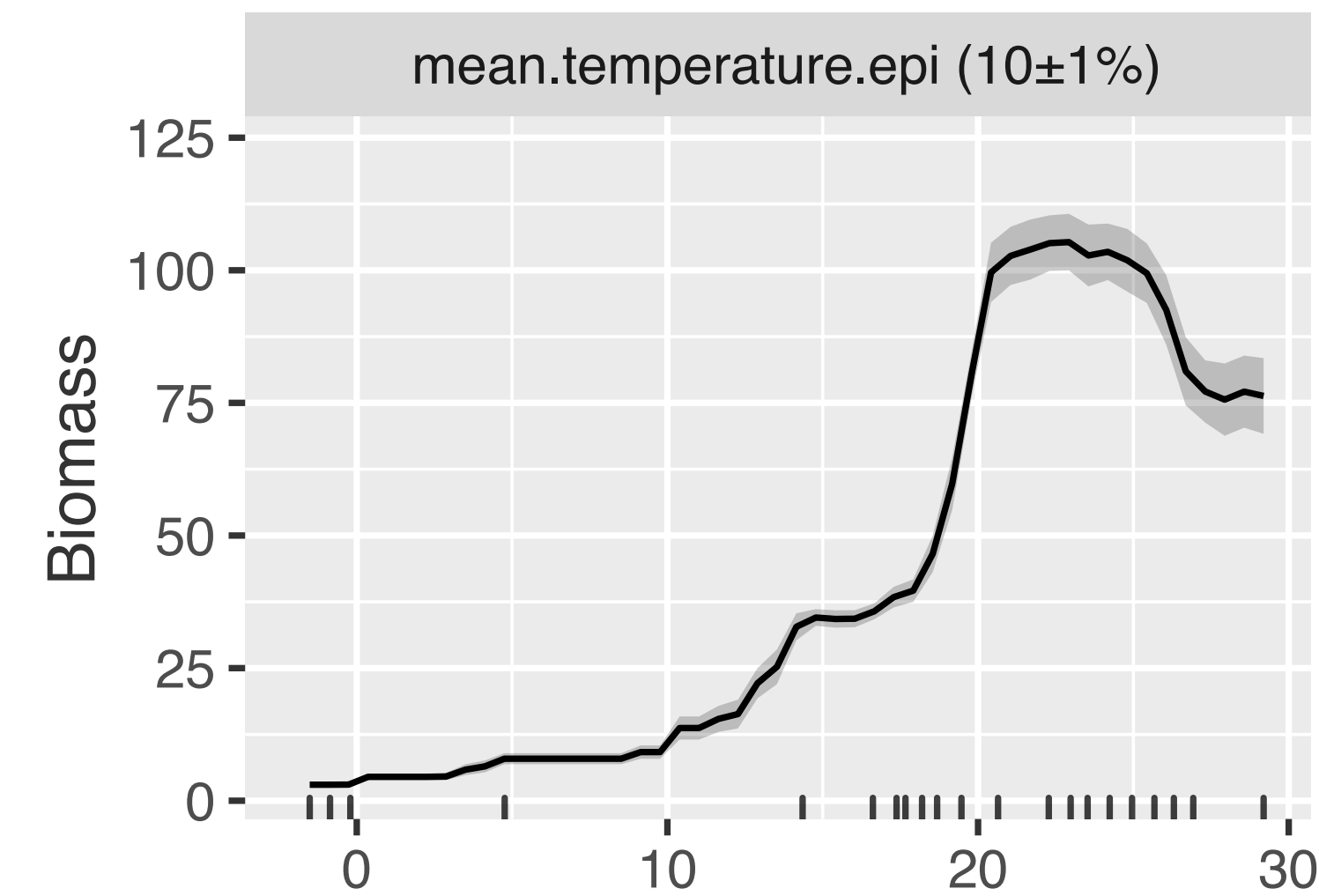
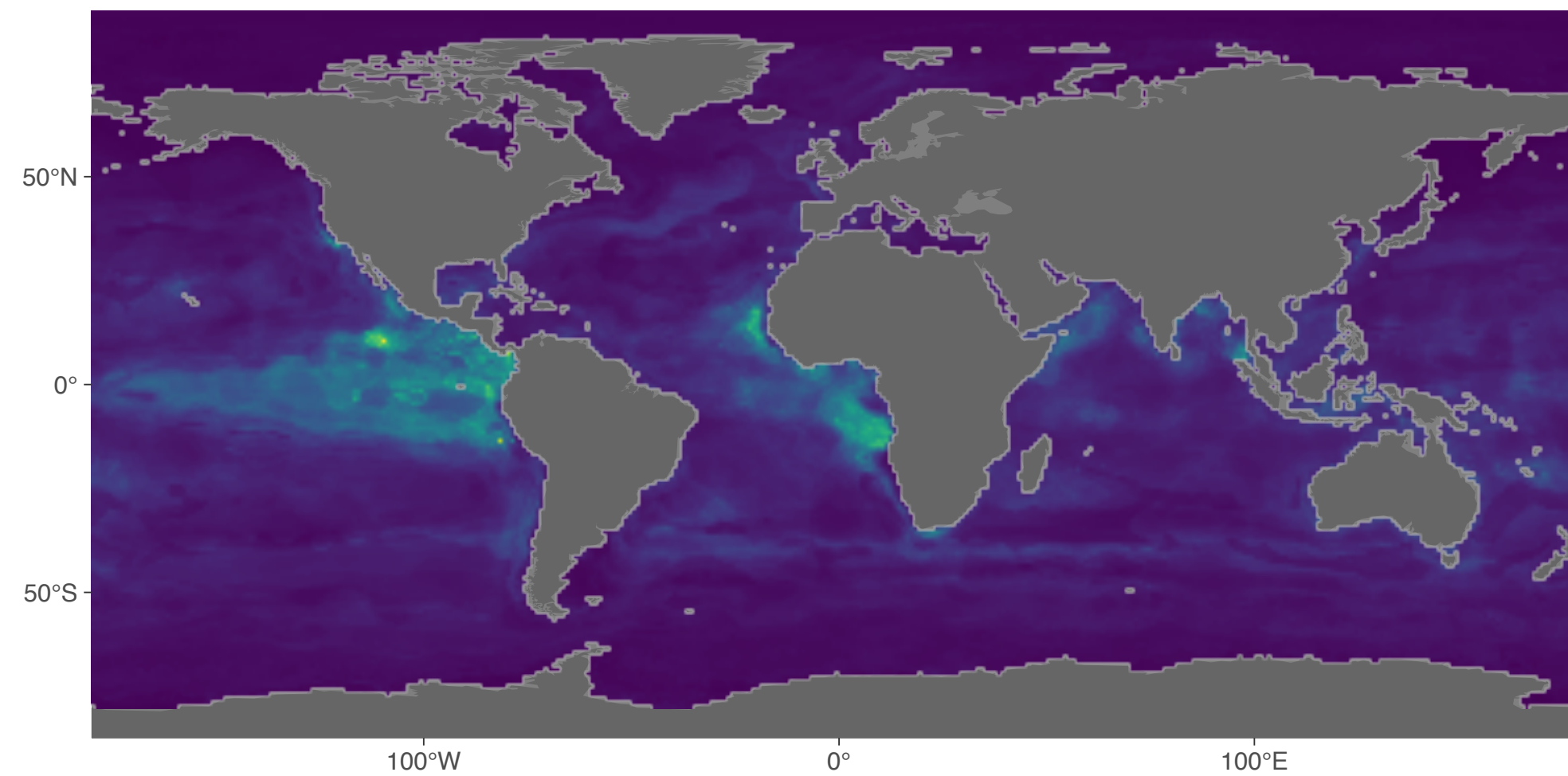
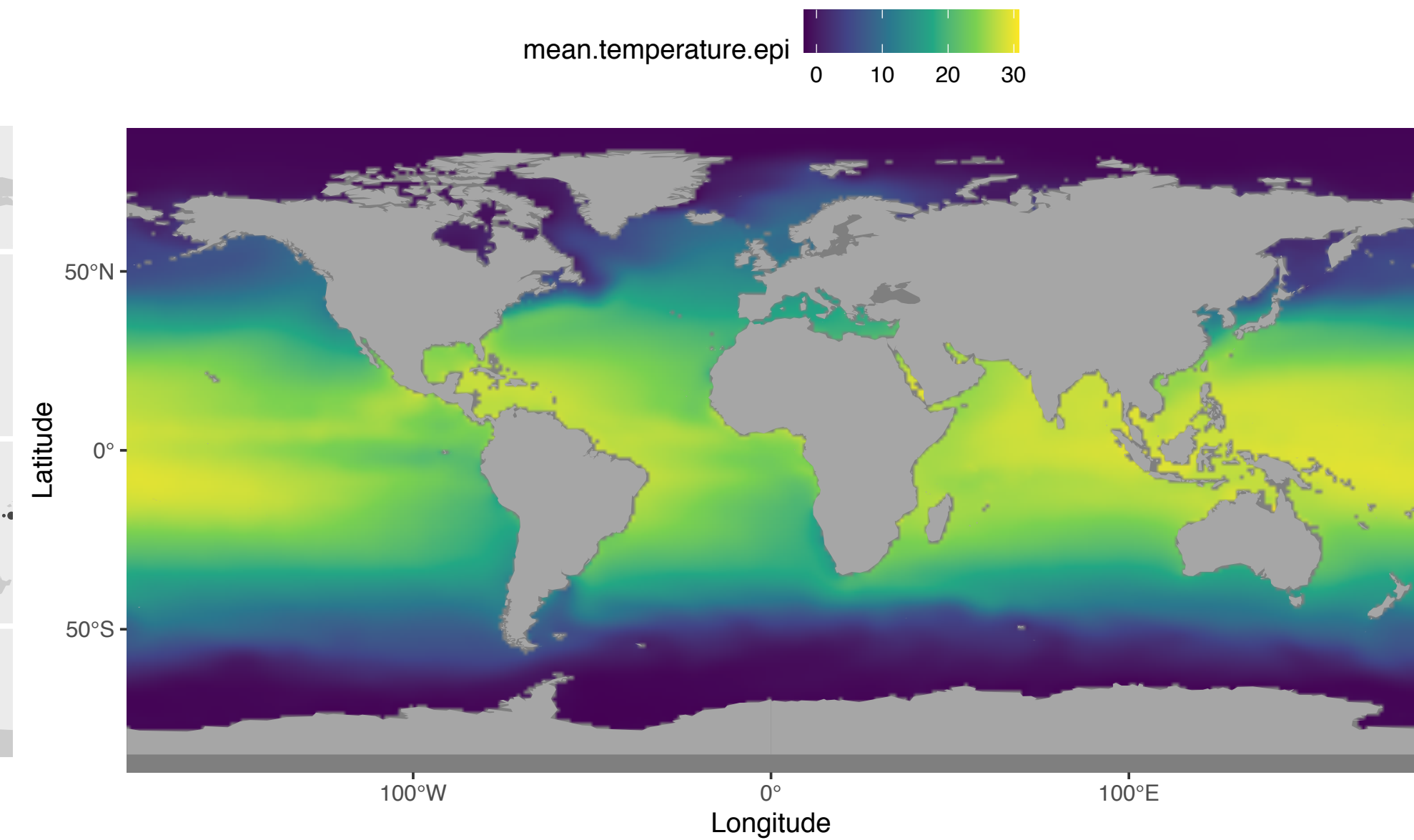
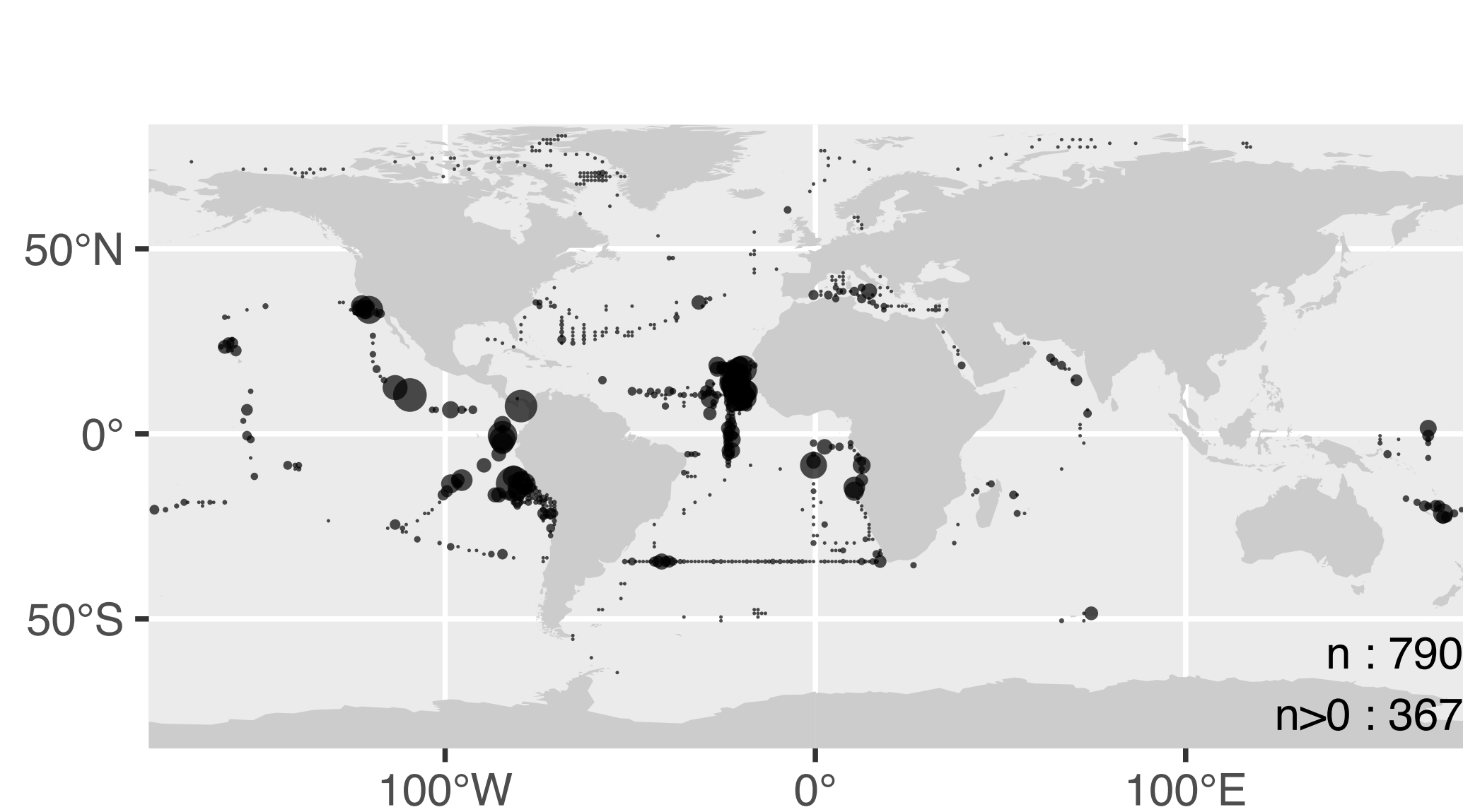
Gradient boosting to predict the basin-scale distribution of organisms



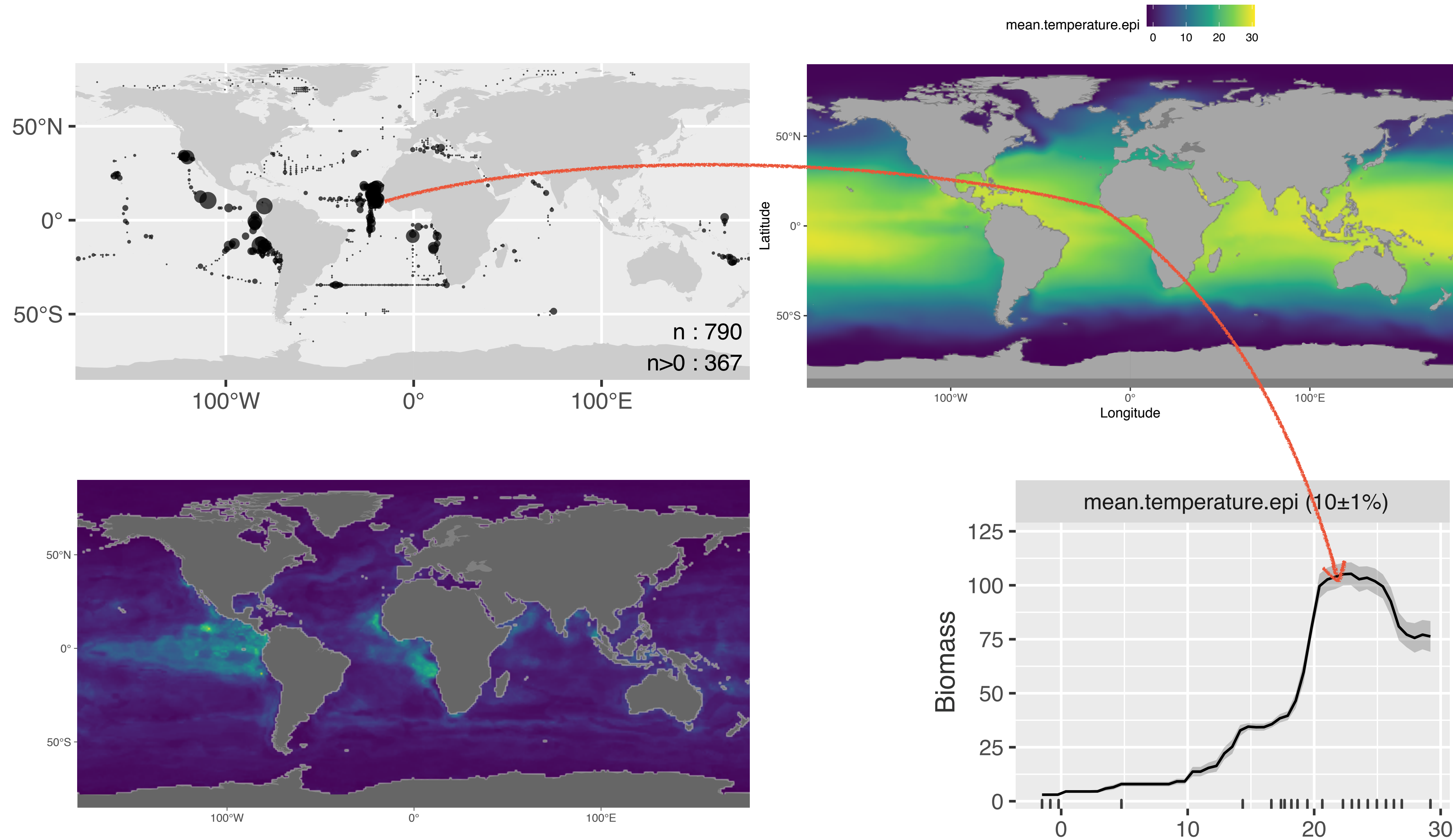
Gradient boosting to predict the basin-scale distribution of organisms



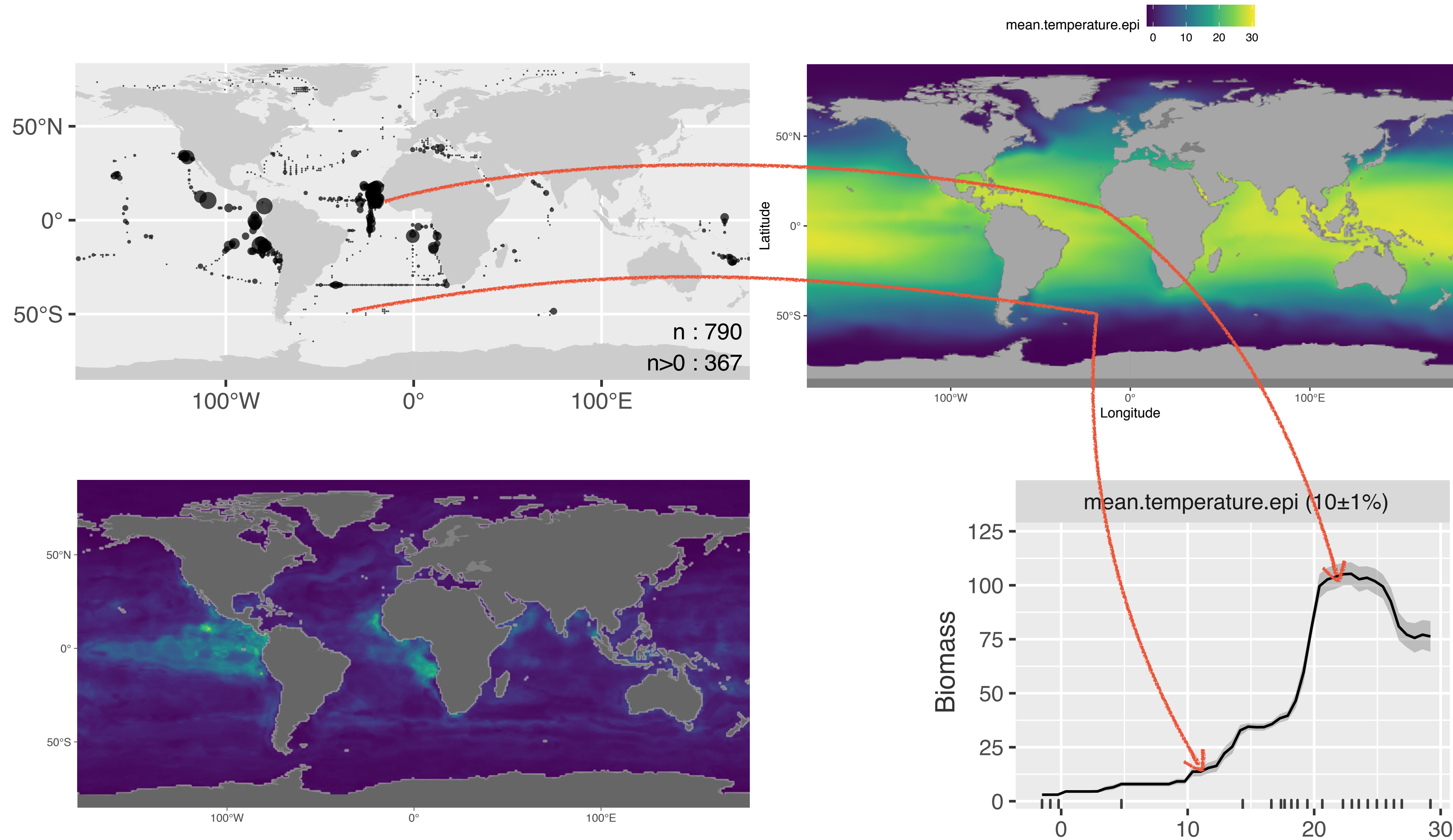
Gradient boosting to predict the basin-scale distribution of organisms



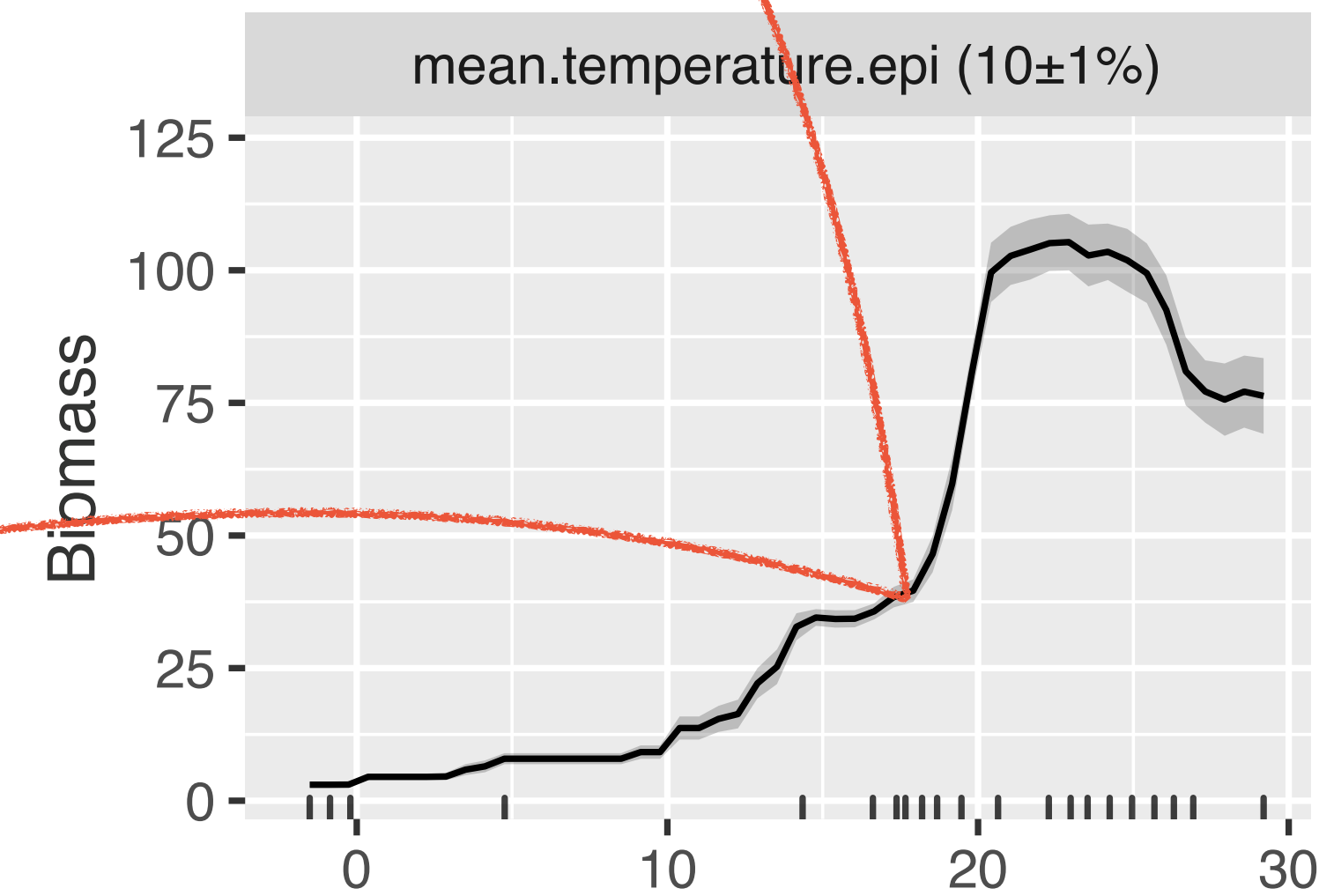
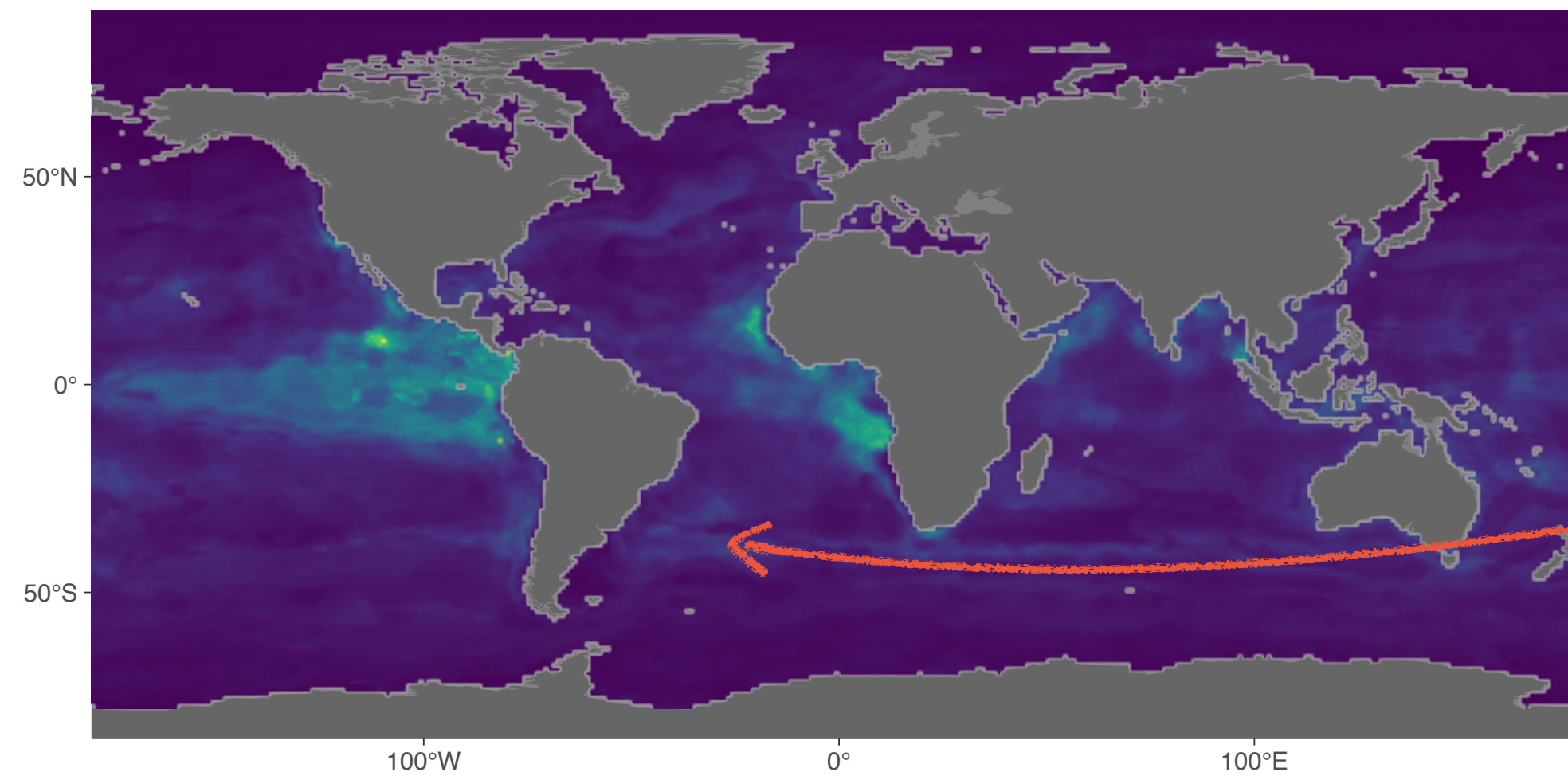
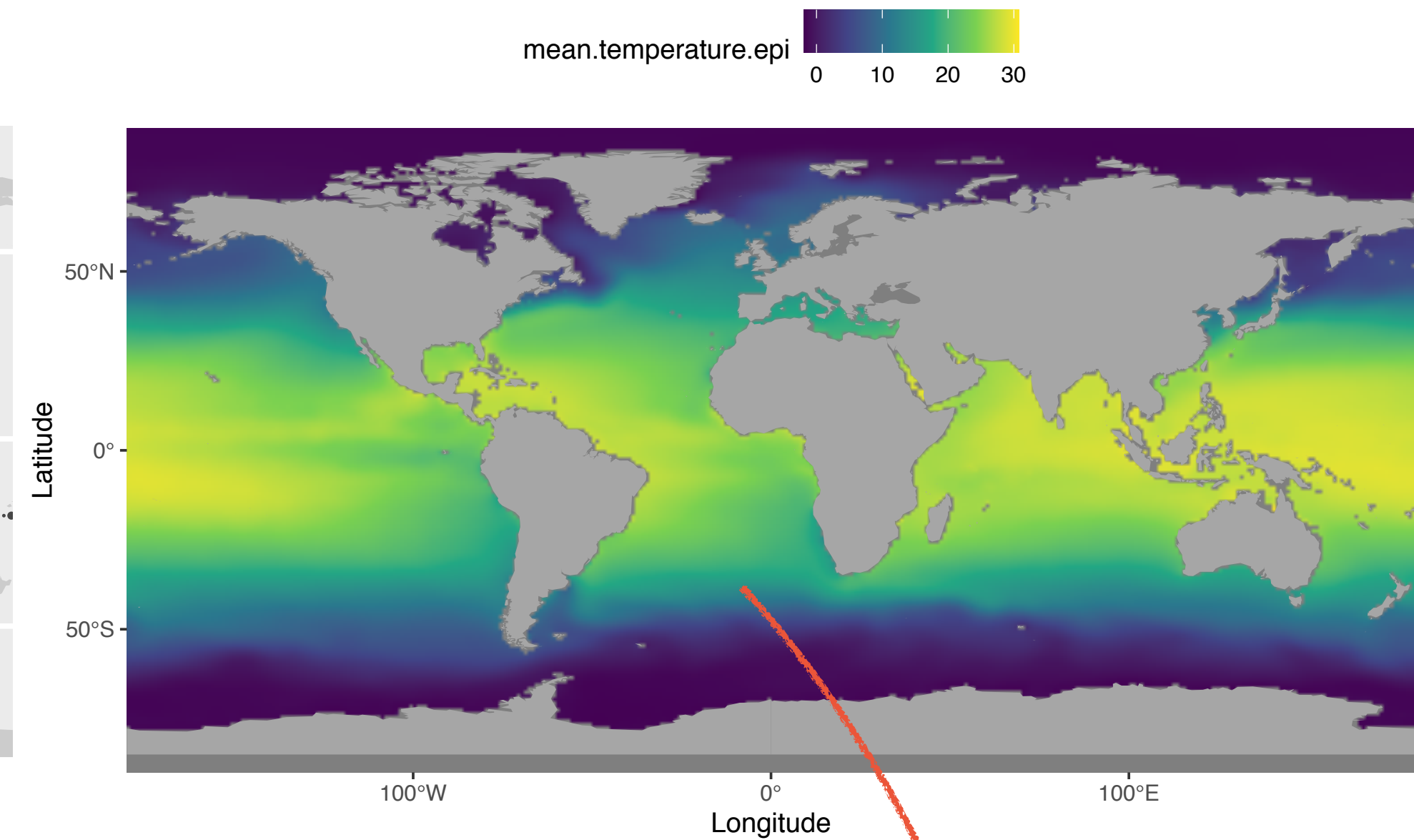
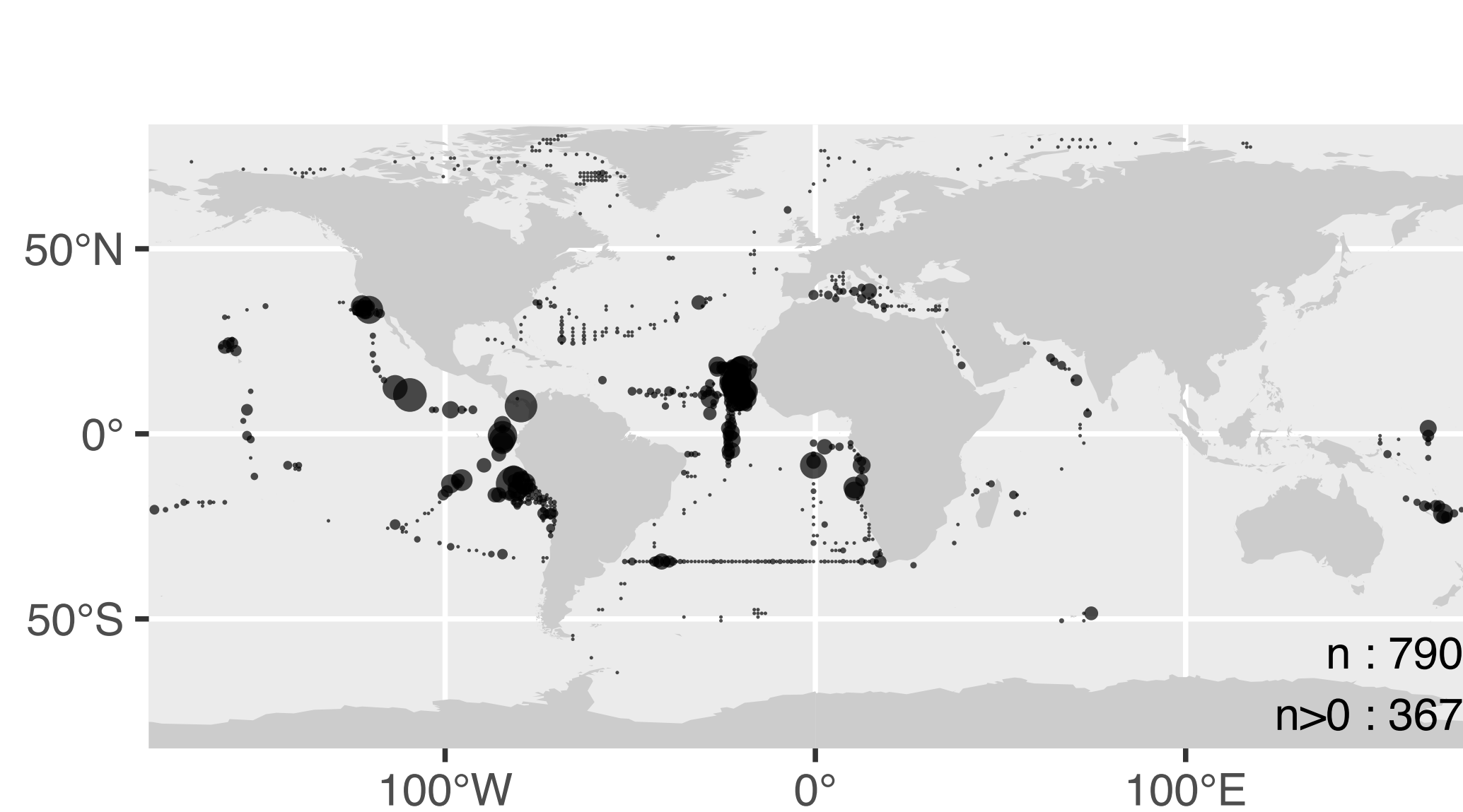
Gradient boosting to predict the basin-scale distribution of organisms



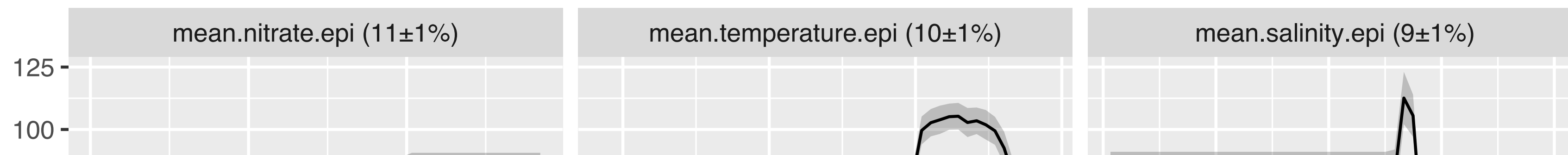
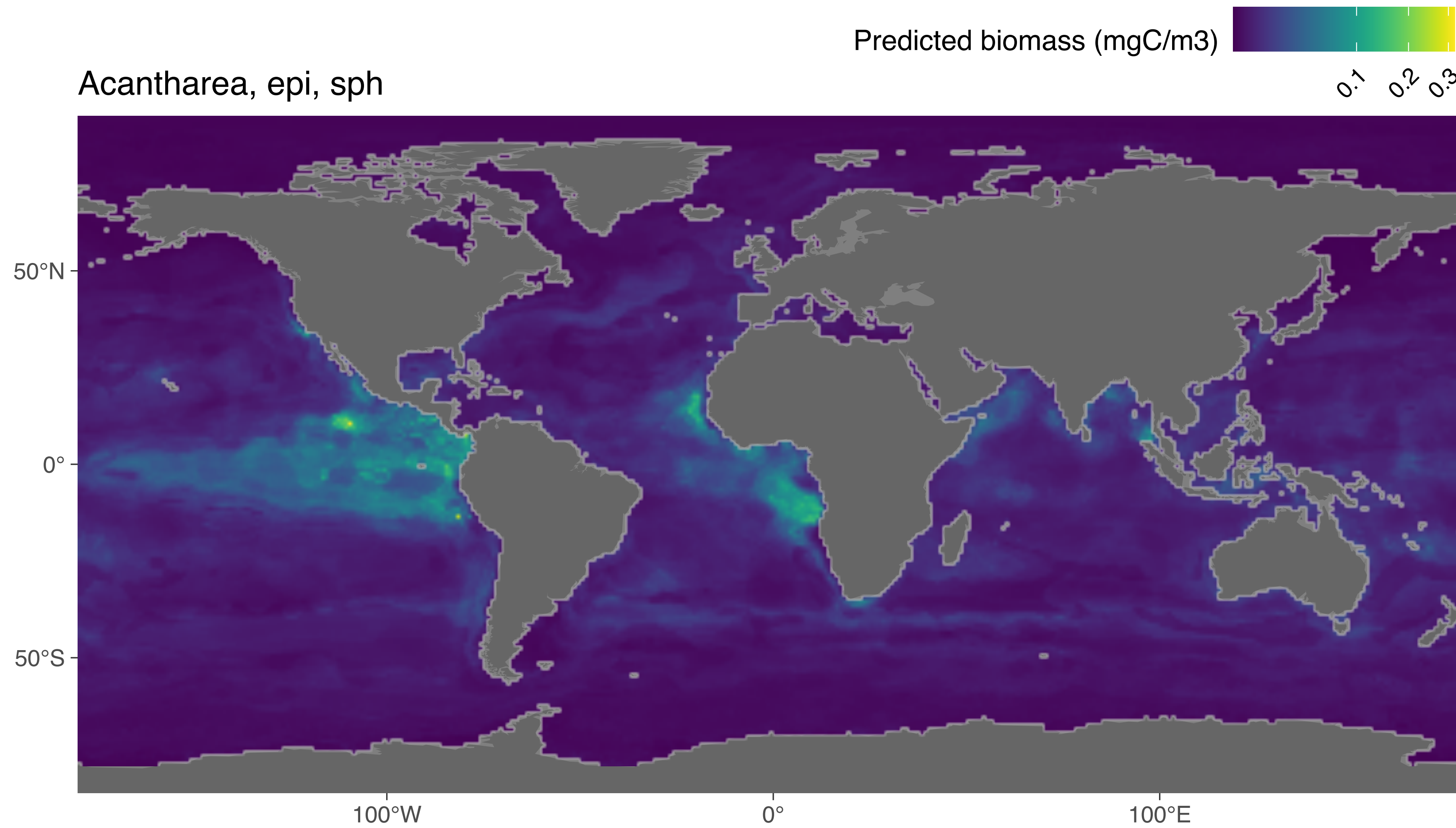
Gradient boosting to predict the basin-scale distribution of organisms



Gradient boosting to predict the basin-scale distribution of organisms



Gradient boosting to predict the basin-scale distribution of organisms



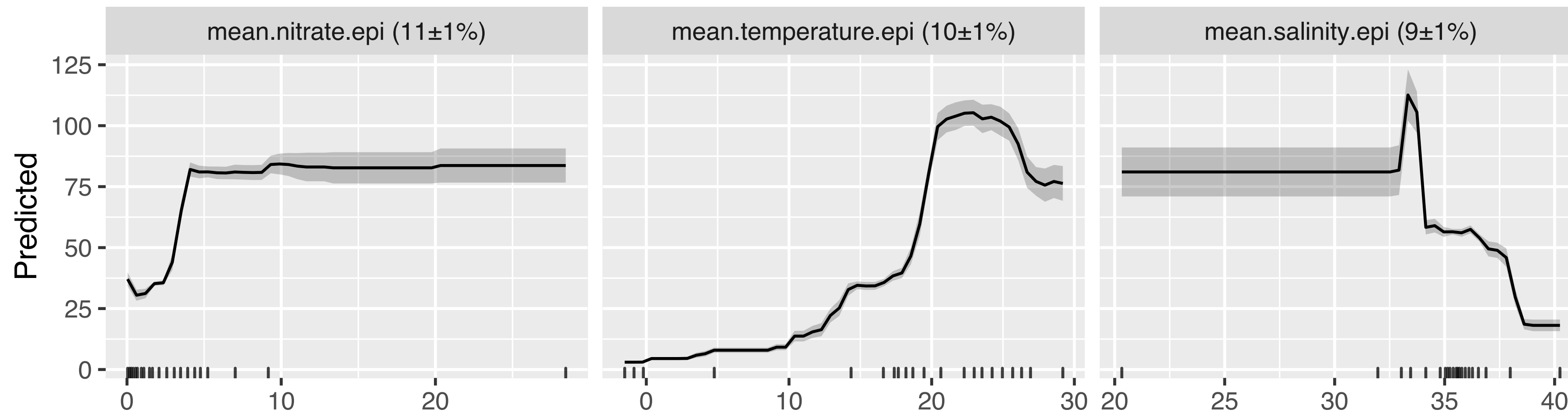
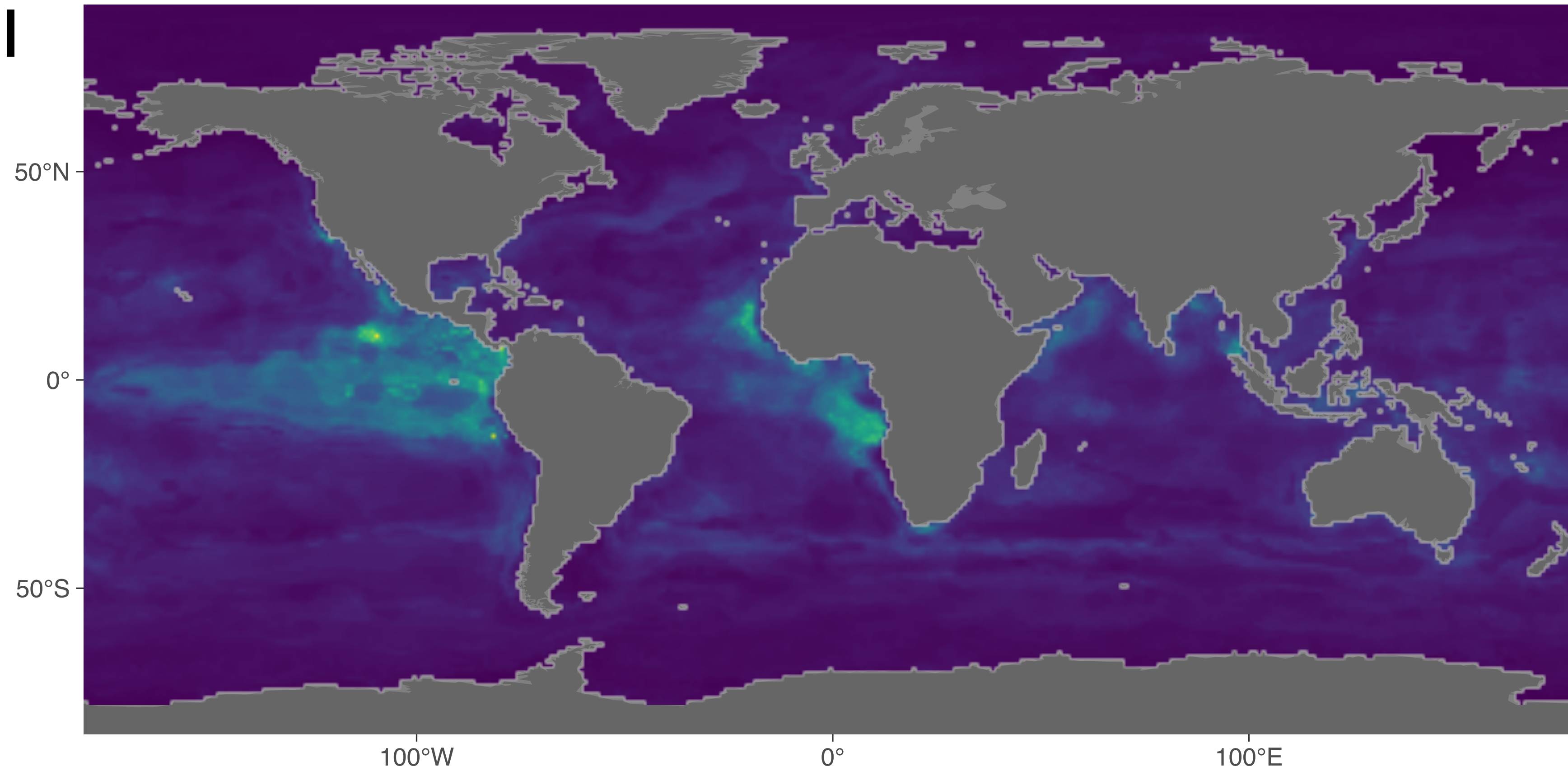
Gradient I

Acantharea, epi, sph

Predicted biomass (mgC/m3)

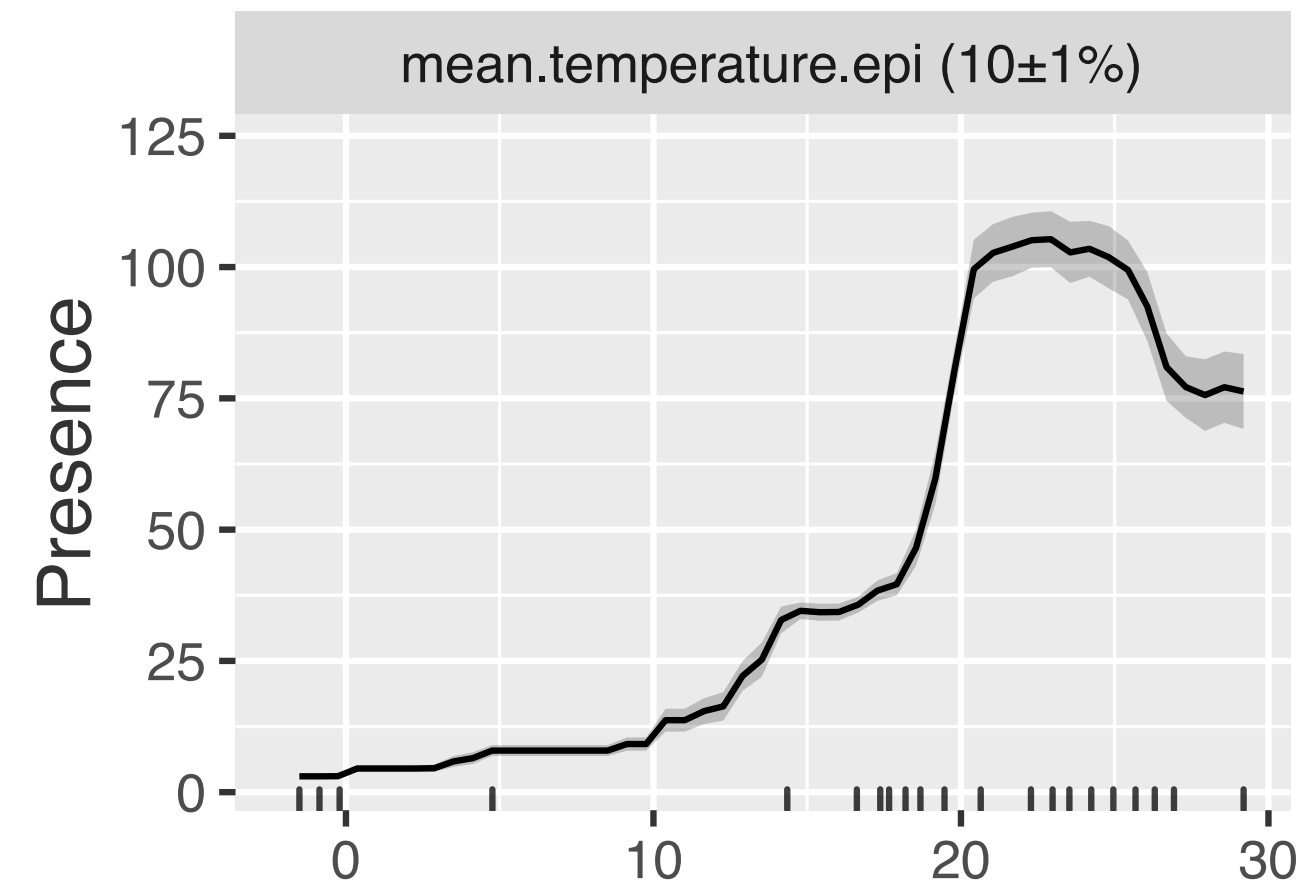
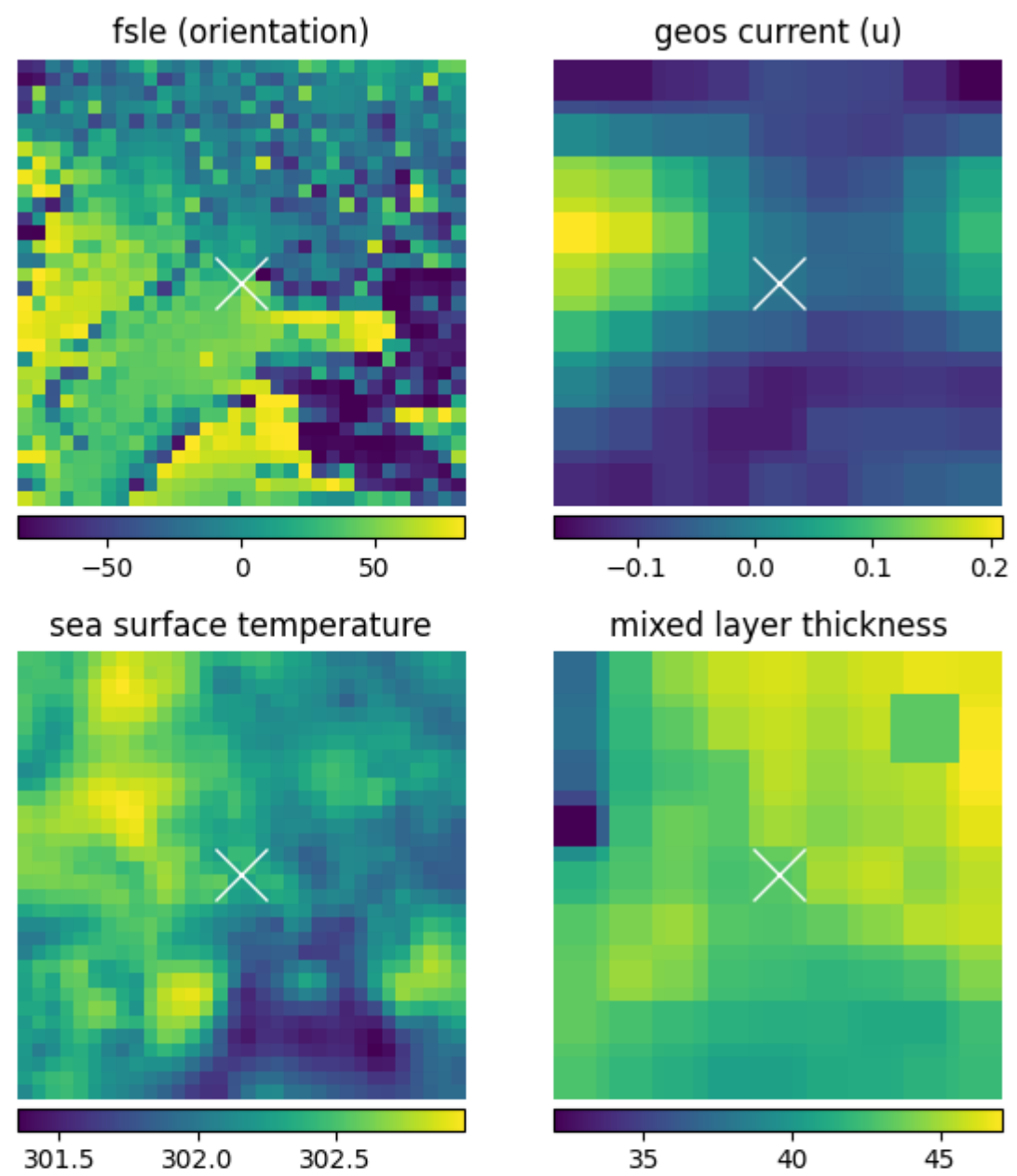
0.1 0.2 0.3

anisms



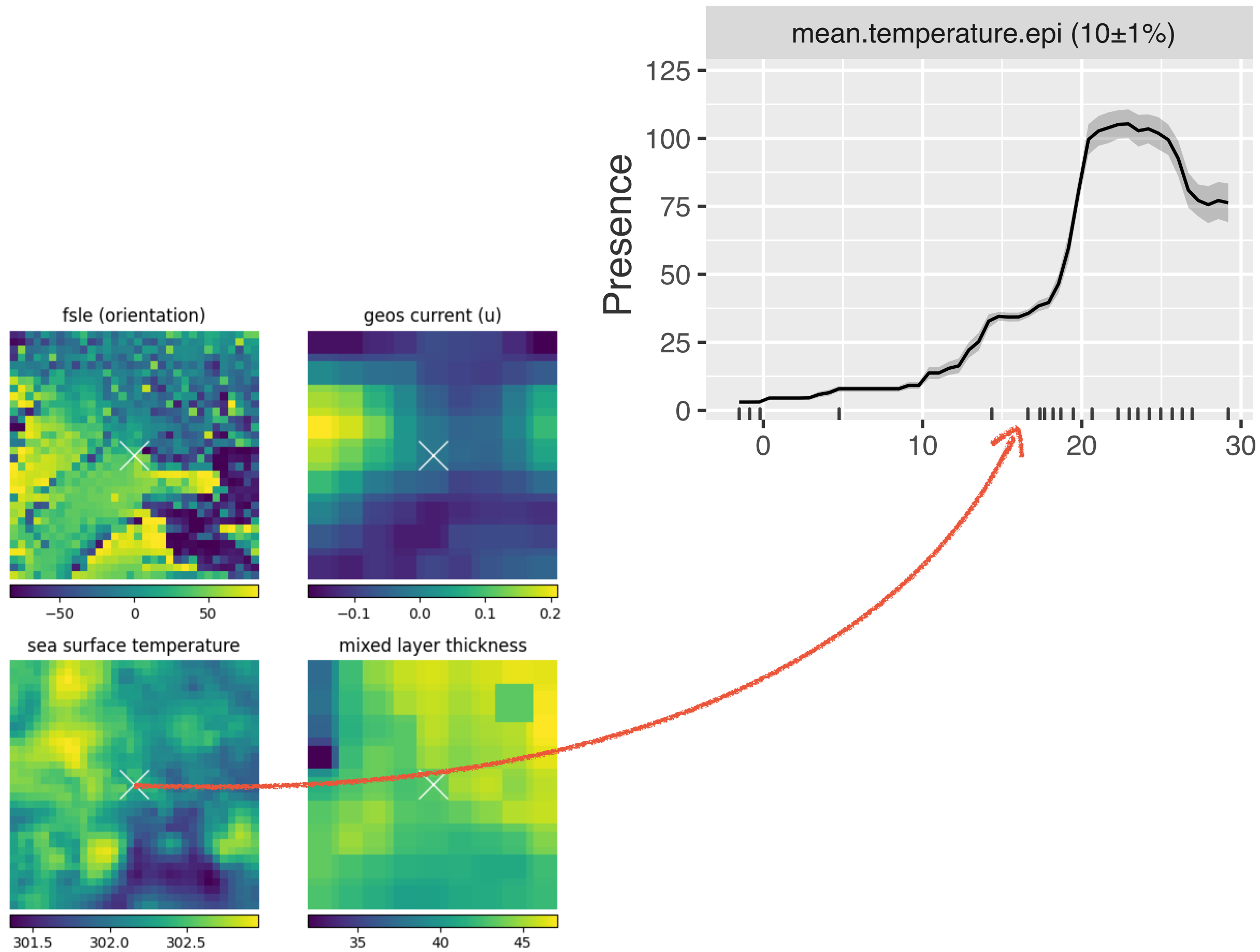
CNNs to take the context into account

Same approach but consider a **full region**, summarised by a CNN, instead of one value



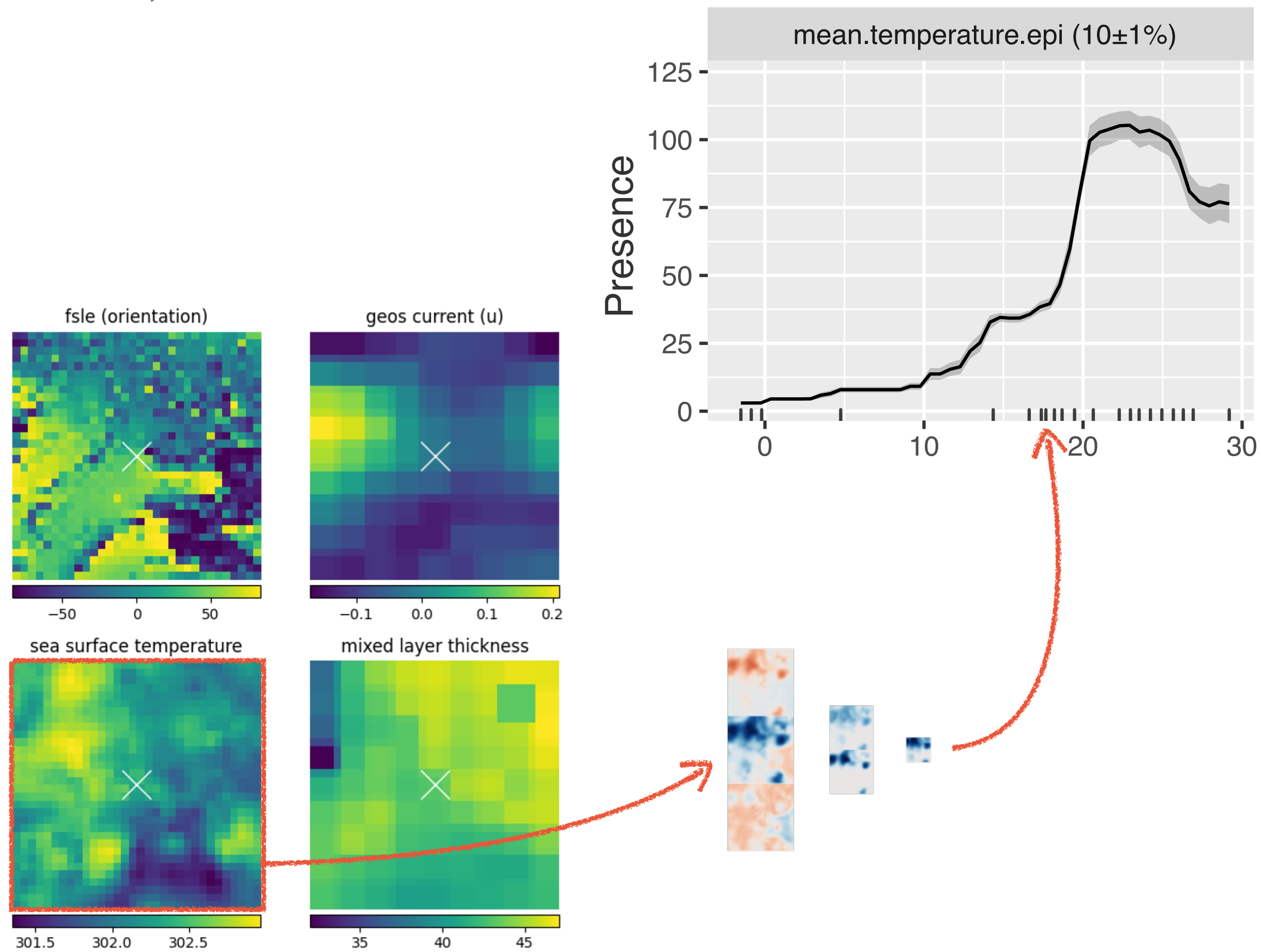
CNNs to take the context into account

Same approach but consider a **full region**, summarised by a CNN, instead of one value



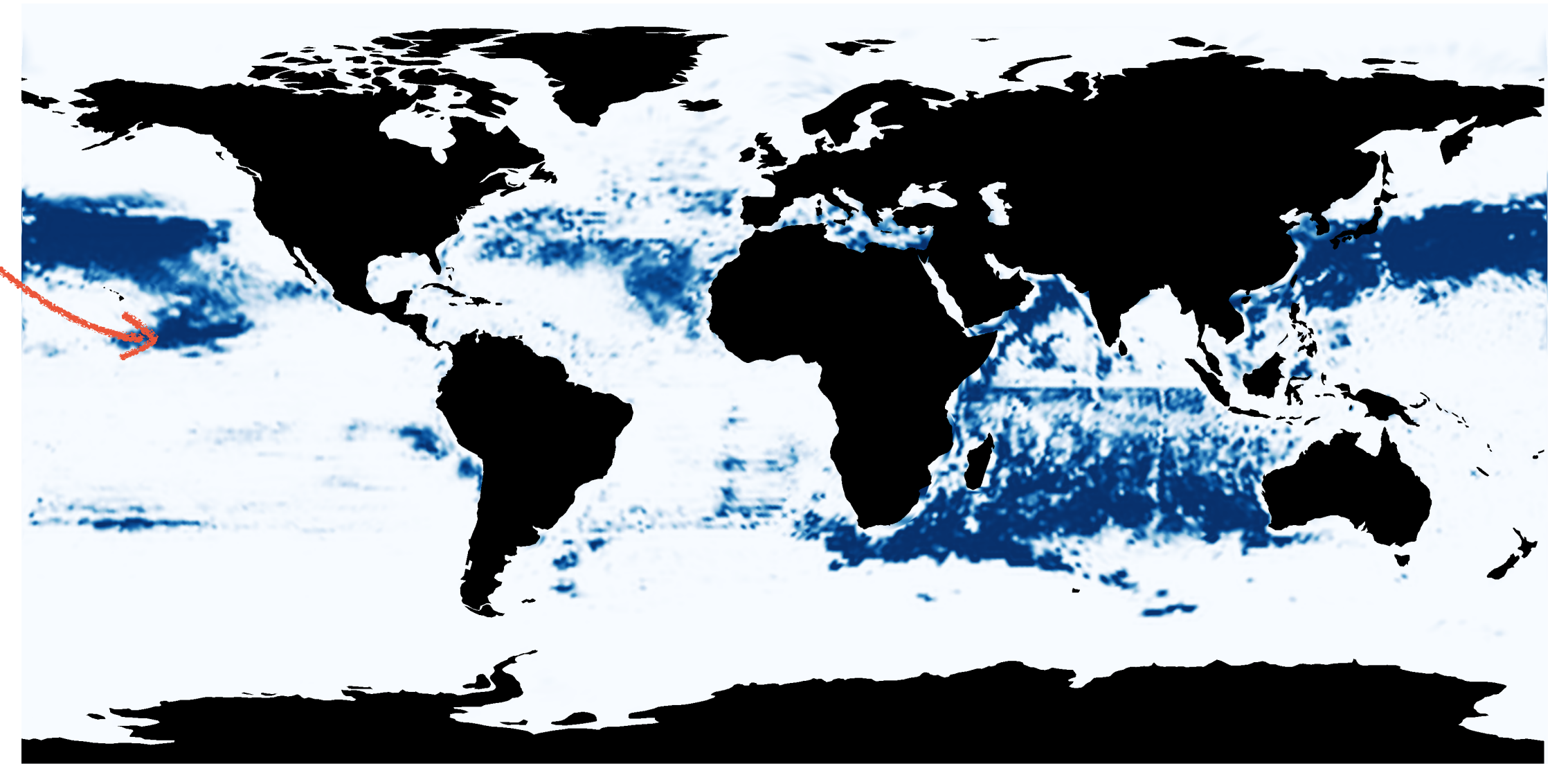
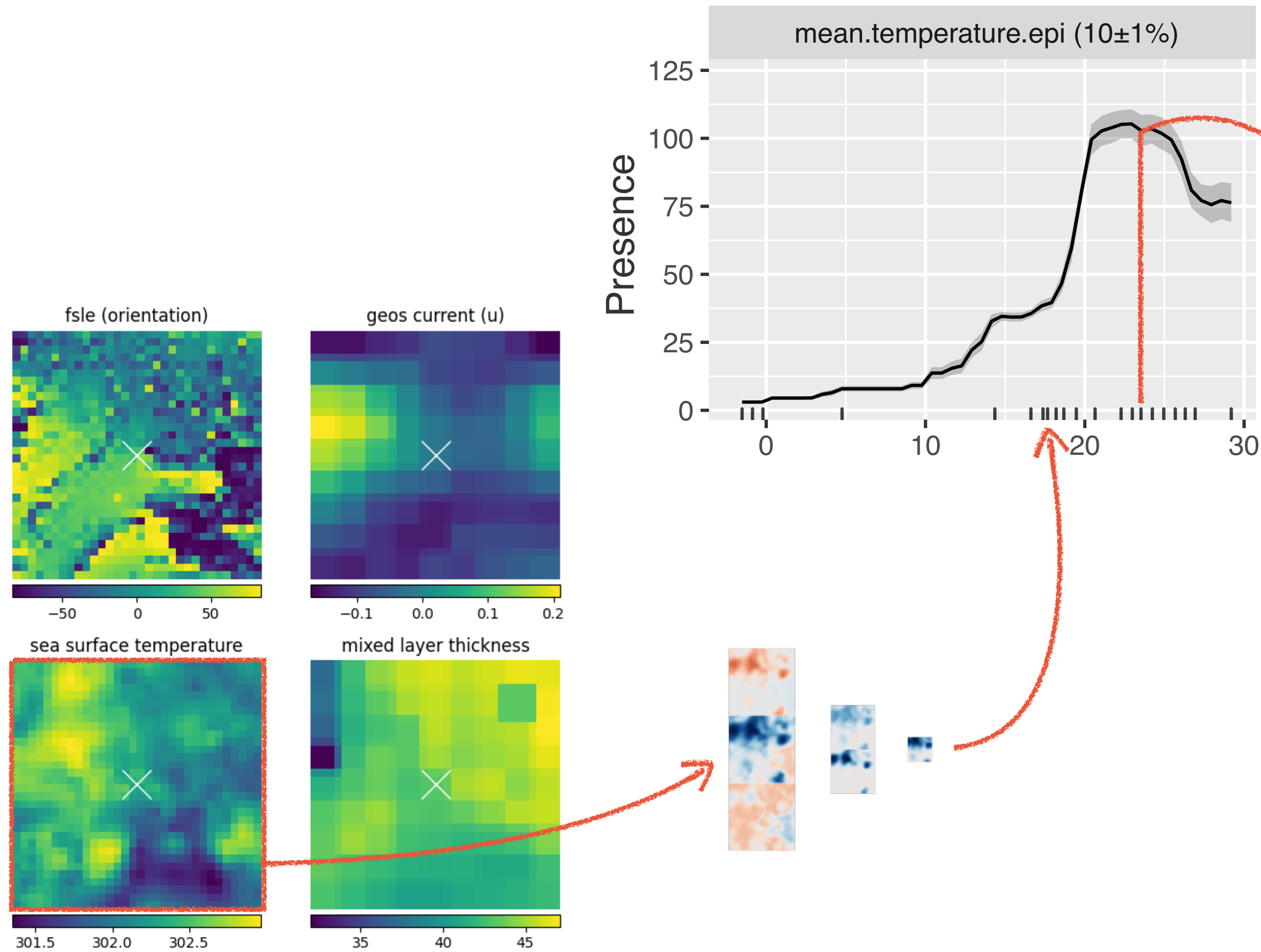
CNNs to take the context into account

Same approach but consider a **full region**, summarised by a CNN, instead of one value



CNNs to take the context into account

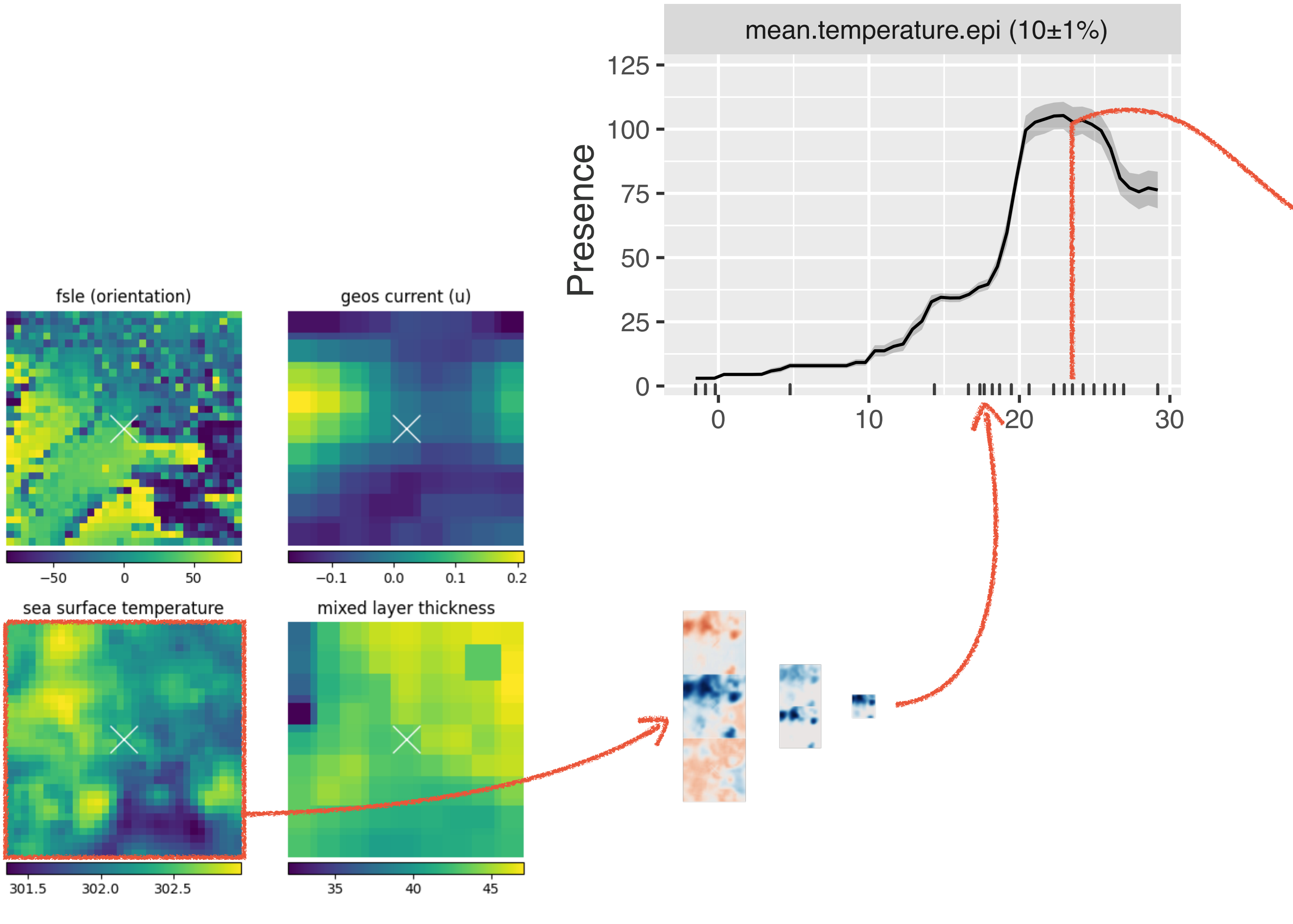
Same approach but consider a **full region**, summarised by a CNN, instead of one value



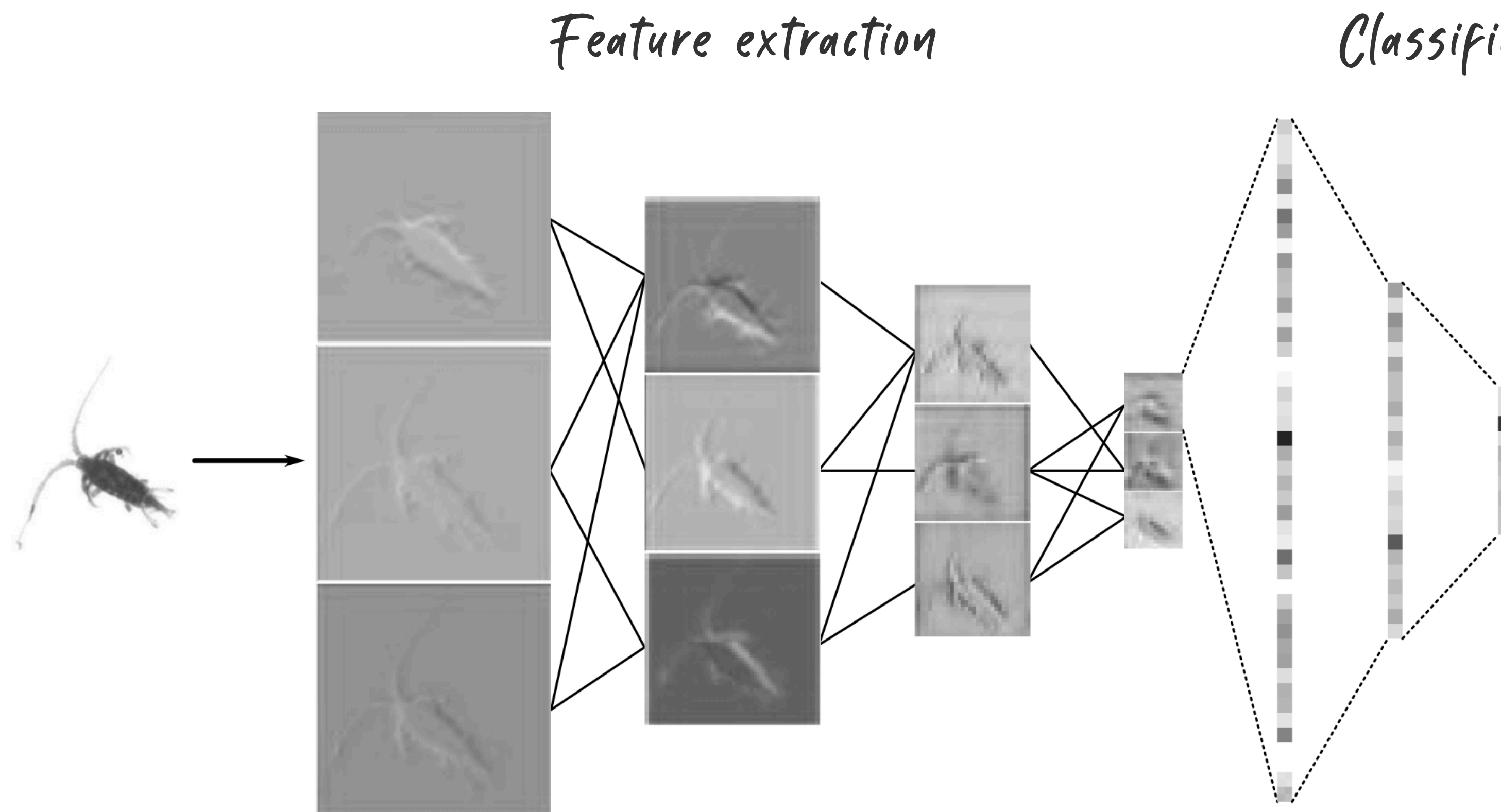
(a) *Caretta caretta*

CNNs to take the context into account

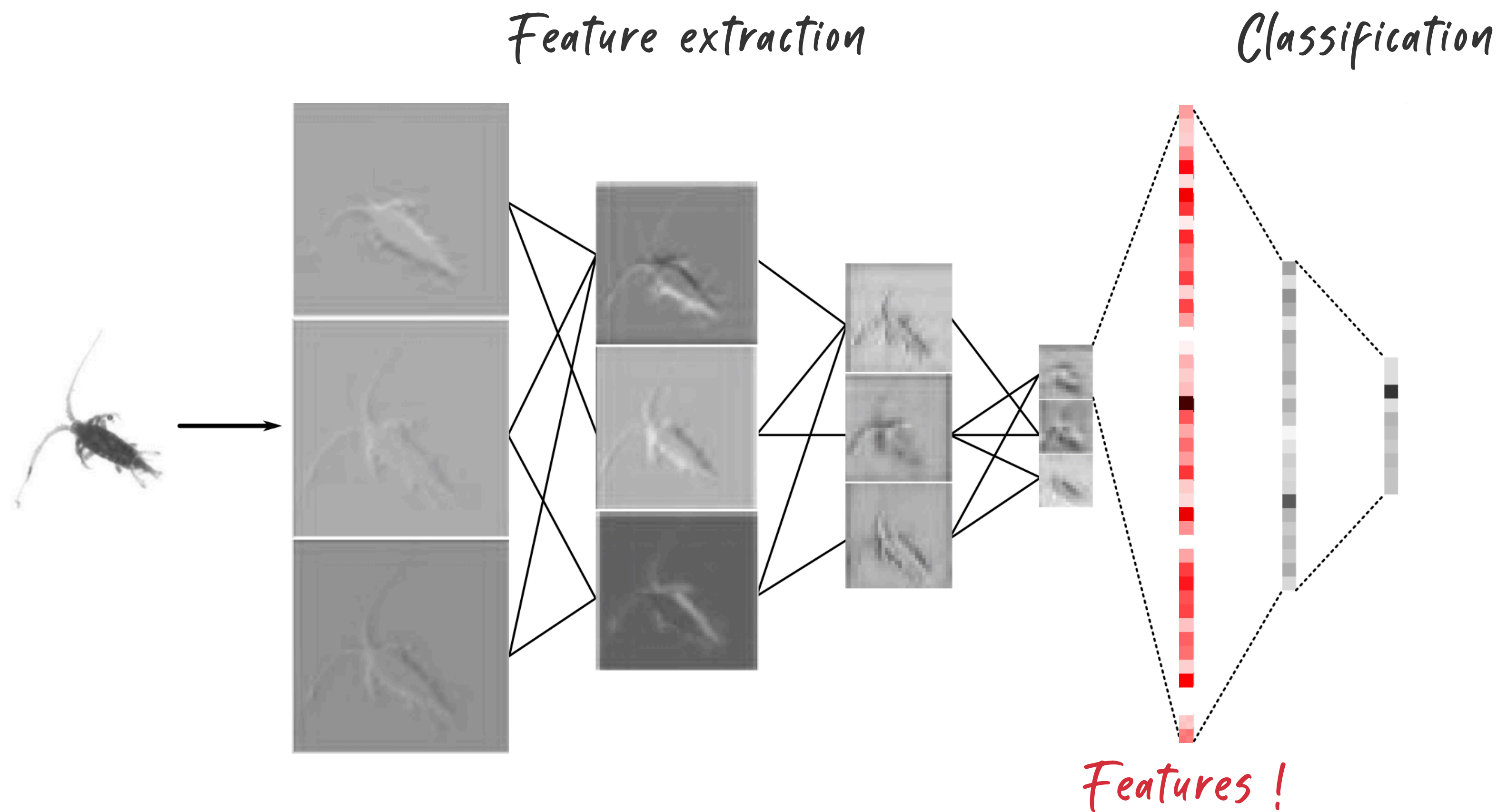
Same approach but consider a **full region**, summarised by a CNN, instead of one value



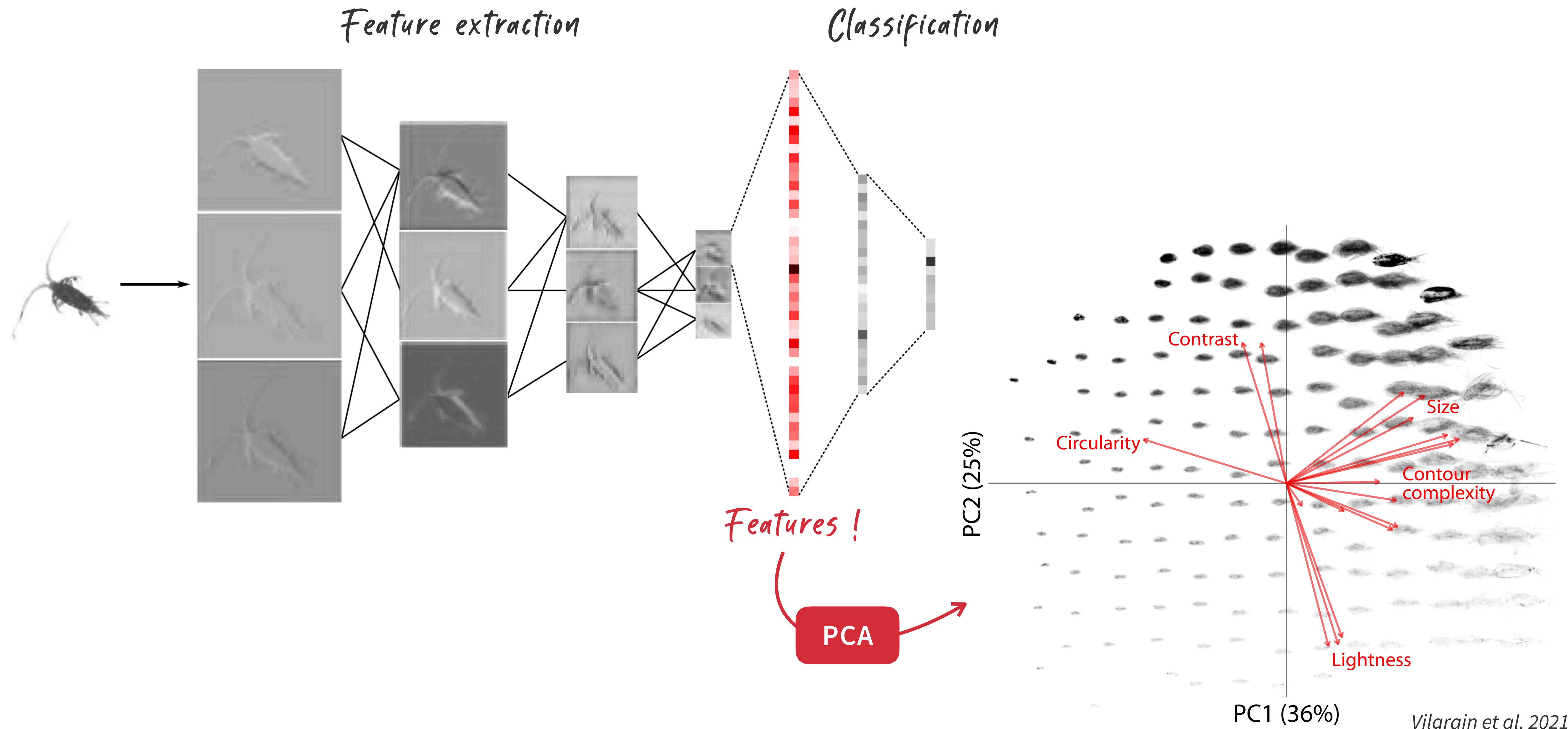
Deep features as objective individual descriptors of morphology



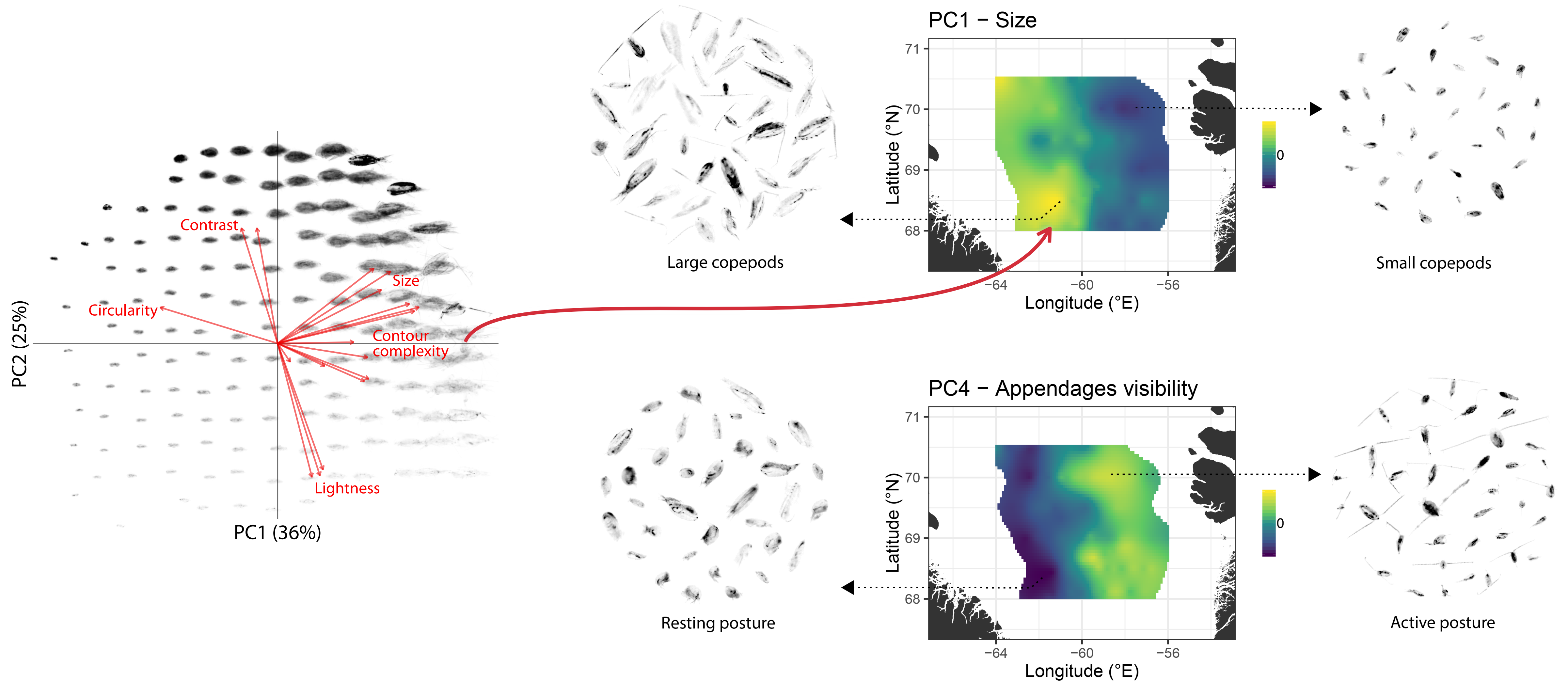
Deep features as objective individual descriptors of morphology



Deep features as objective individual descriptors of morphology



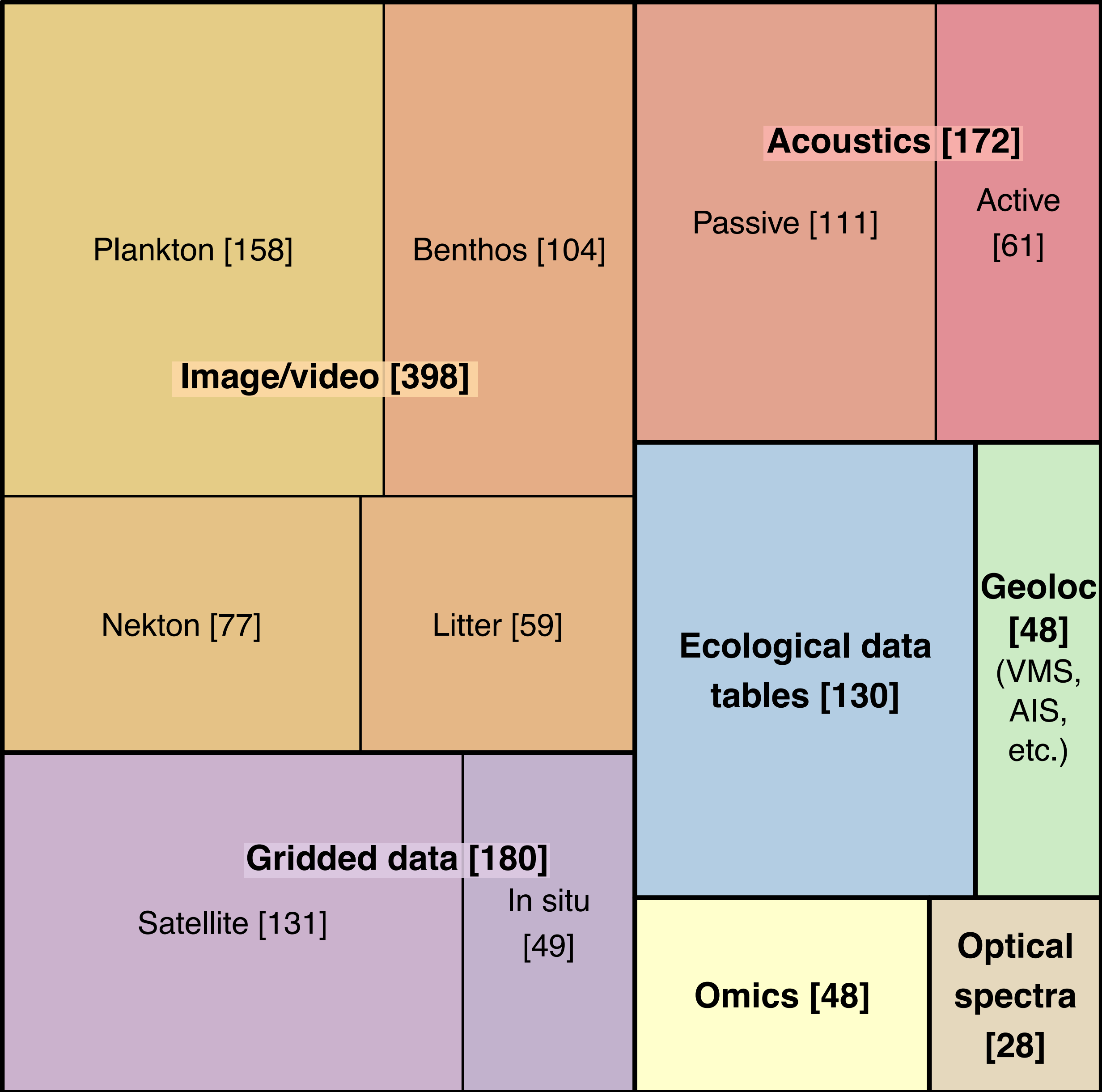
Deep features as objective individual descriptors of morphology



Overall conclusions

What? When? Who?





Data types

Based on the **tags** of the ~1000 papers in the library

45% images + 15% satellite images = **60% images**

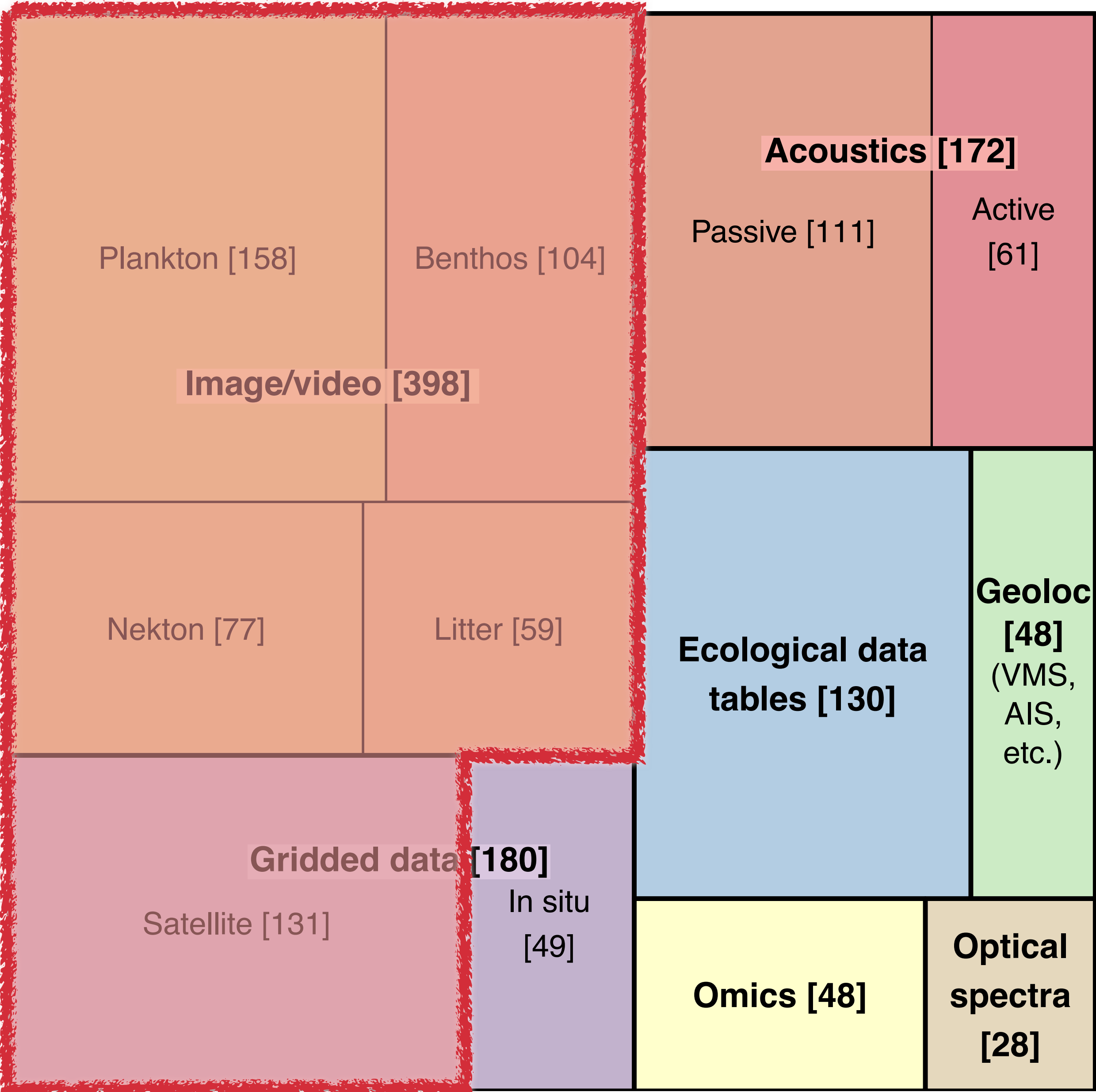
Why such a success?

Data types

Based on the **tags** of the ~1000 papers in the library

45% images + 15% satellite images = **60% images**

Why such a success?

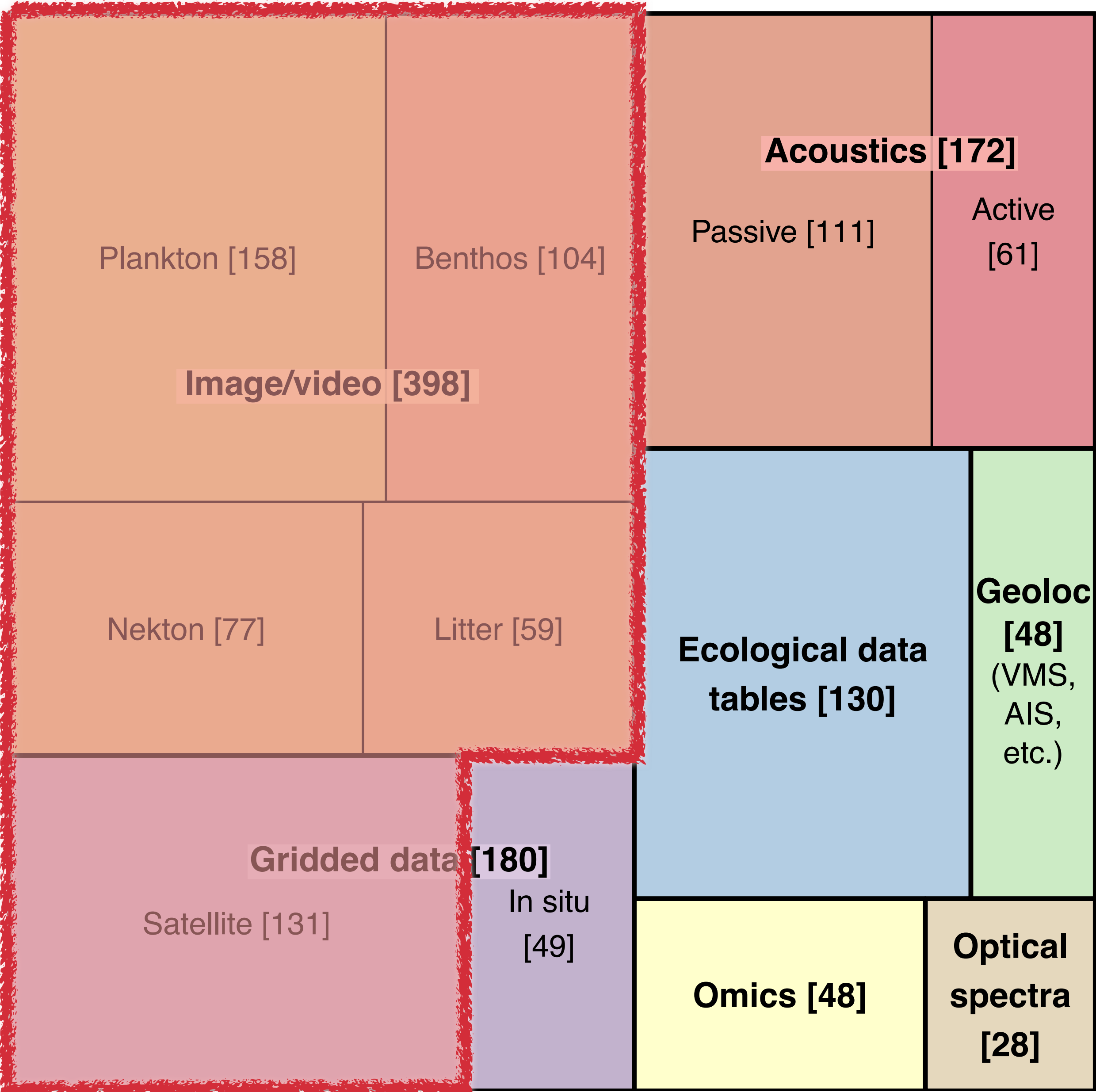


Data types

Based on the **tags** of the ~1000 papers in the library

45% images + 15% satellite images = **60% images**

Why such a success?



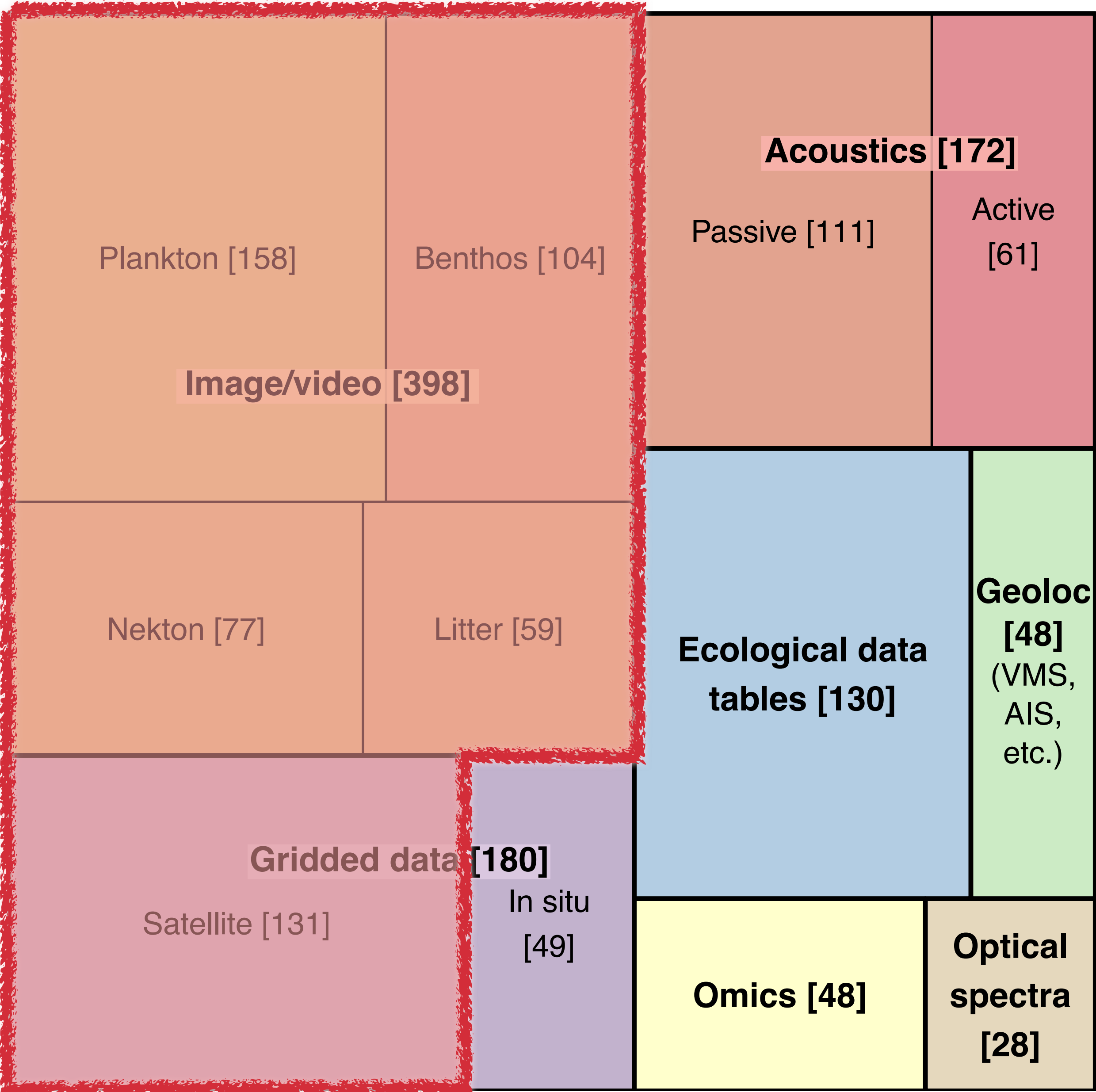
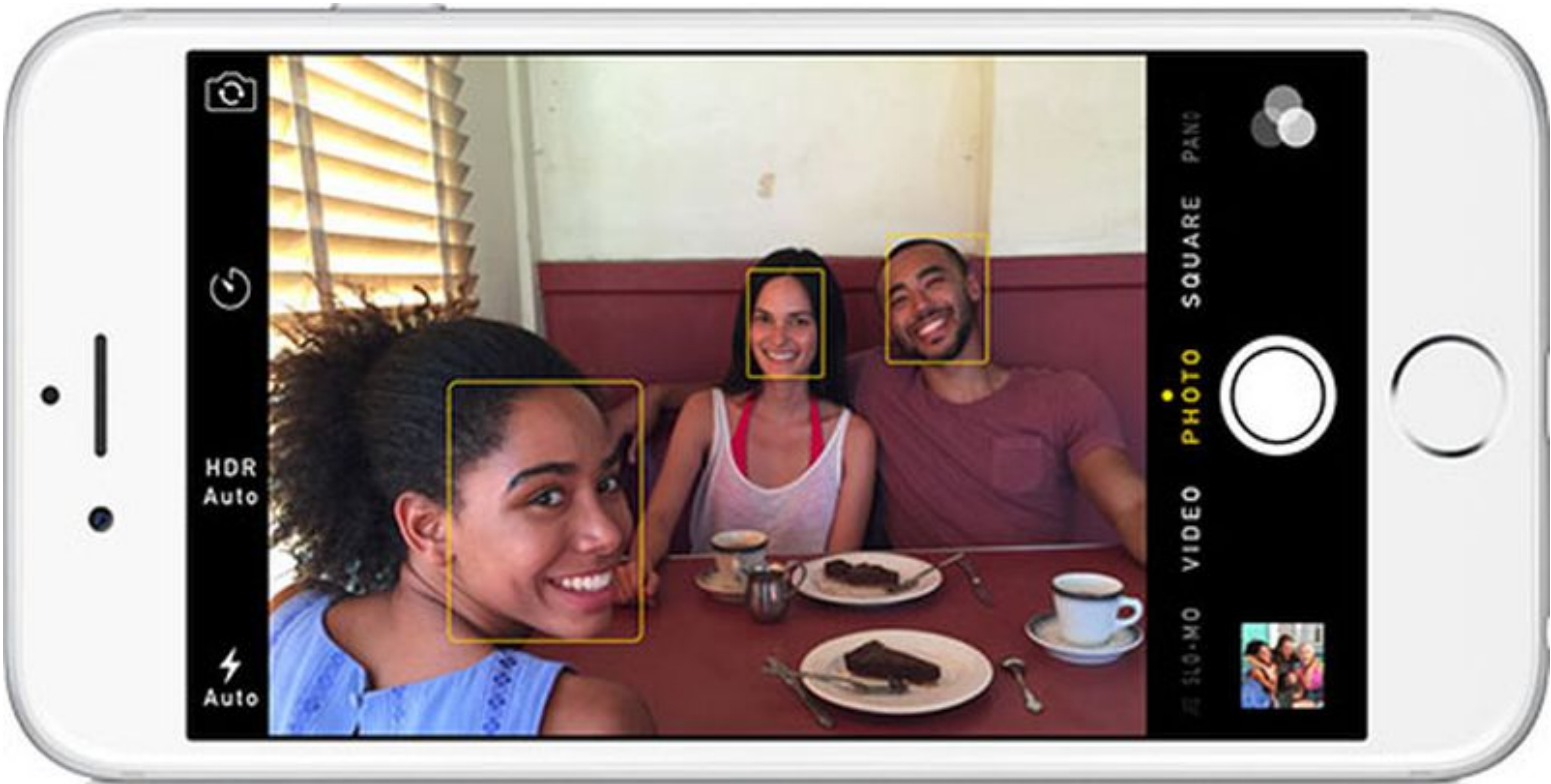
Data types

Based on the **tags** of the ~1000 papers in the library

45% images + 15% satellite images = **60% images**

Why such a success?

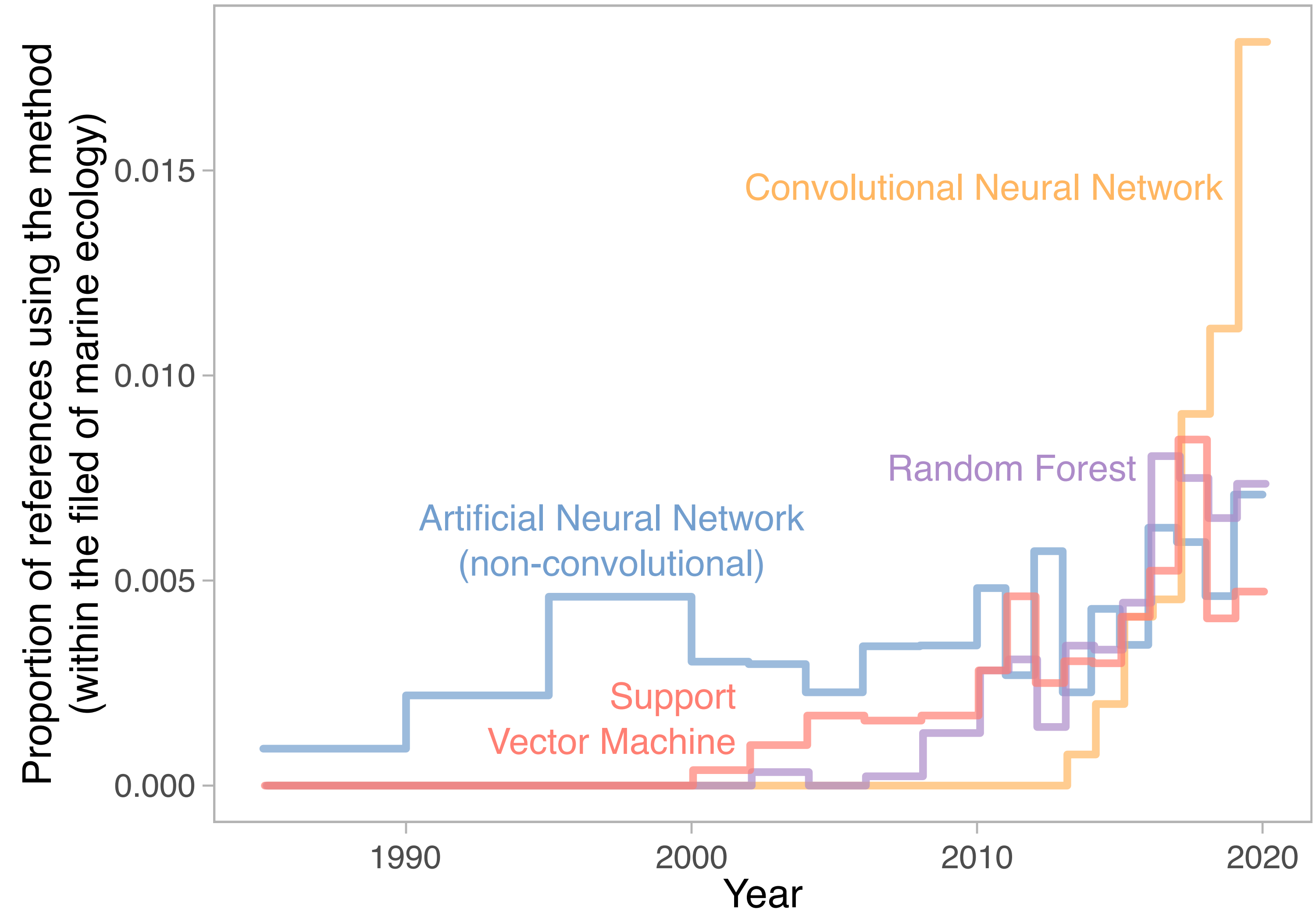
Because images are common in everyday life



Techniques

Machine learning as a whole is becoming more **popular** in marine ecology

As expected, **CNNs explode** in popularity in recent years

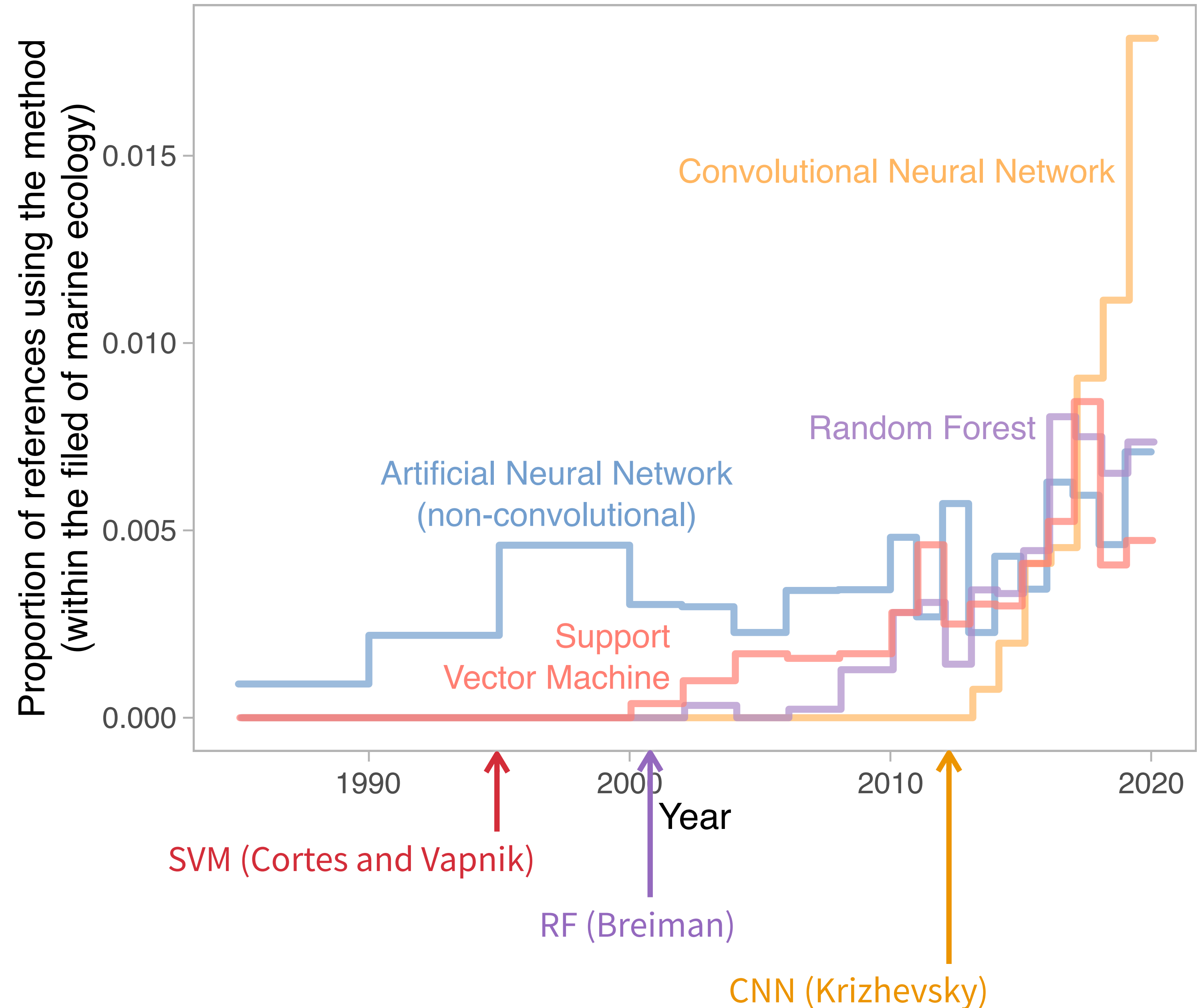


Techniques

Machine learning as a whole is becoming more **popular** in marine ecology

As expected, **CNNs explode** in popularity in recent years

Quick adoption of novel techniques in marine ecology

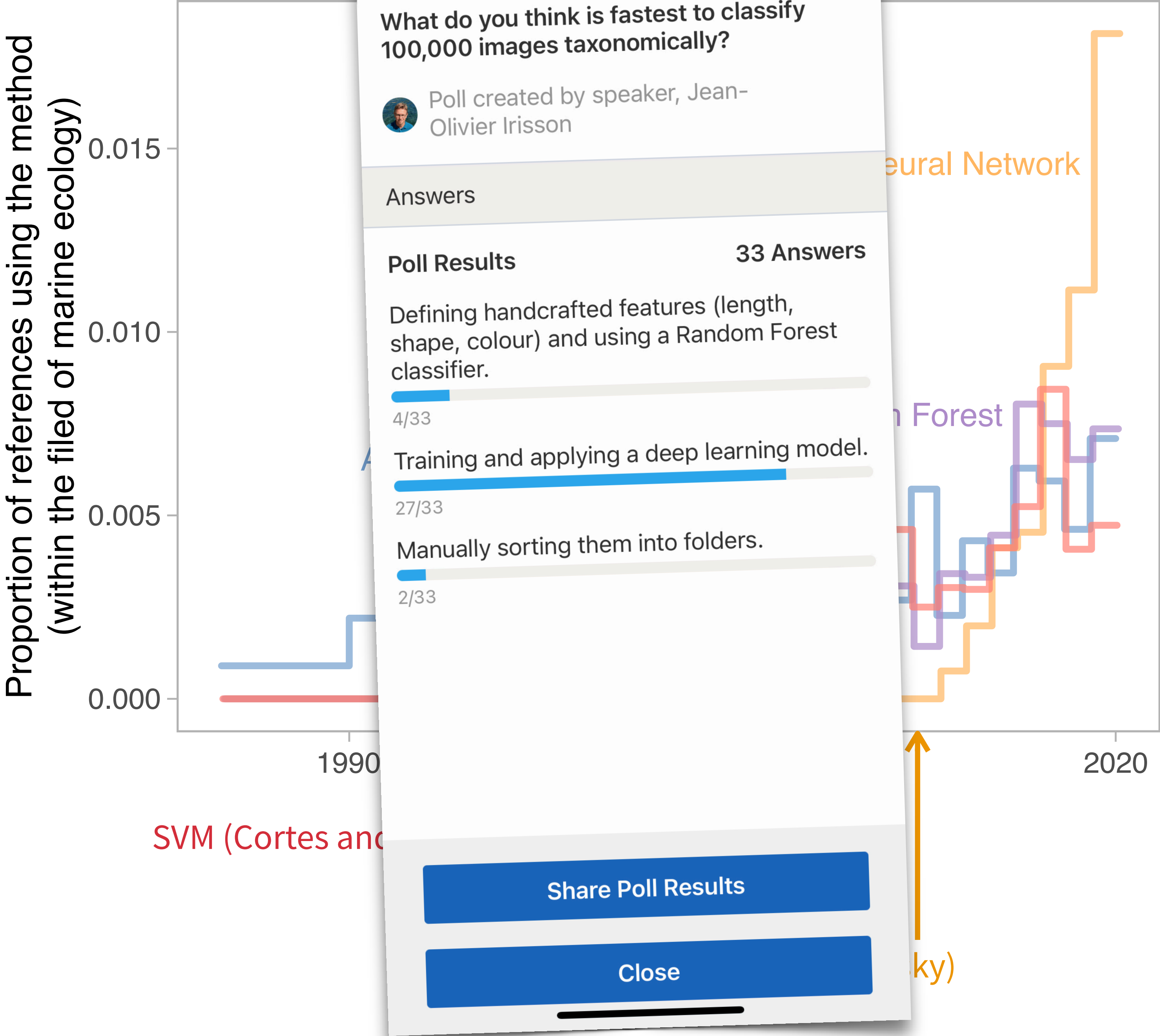


Techniques

Machine learning as a whole is becoming more **popular** in marine ecology

As expected, **CNNs explode** in popularity in recent years

Quick adoption of novel techniques in marine ecology



Limitations of machine learning applications in marine ecology

Feeling that machine/deep learning is **less explainable** than traditional statistics



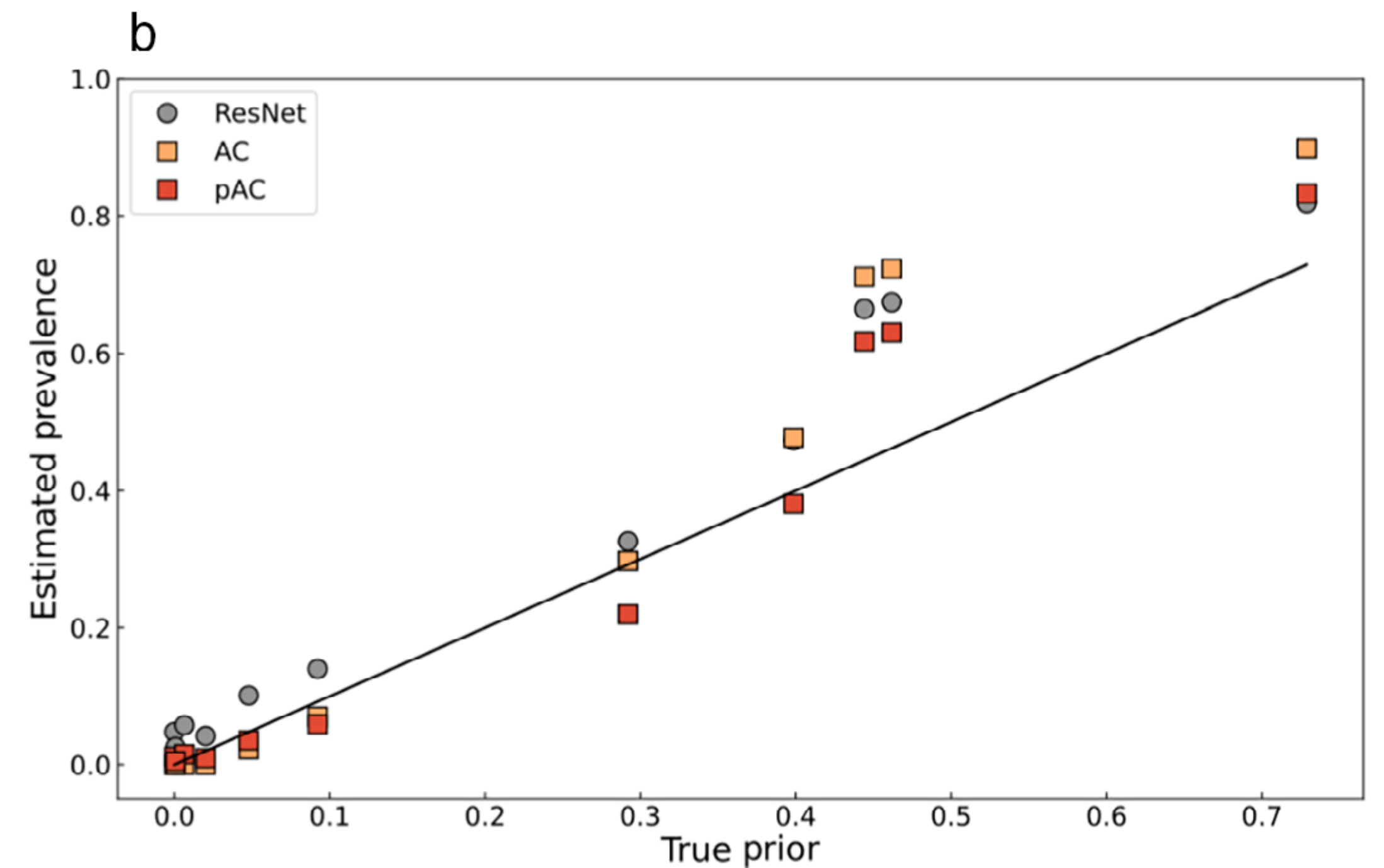
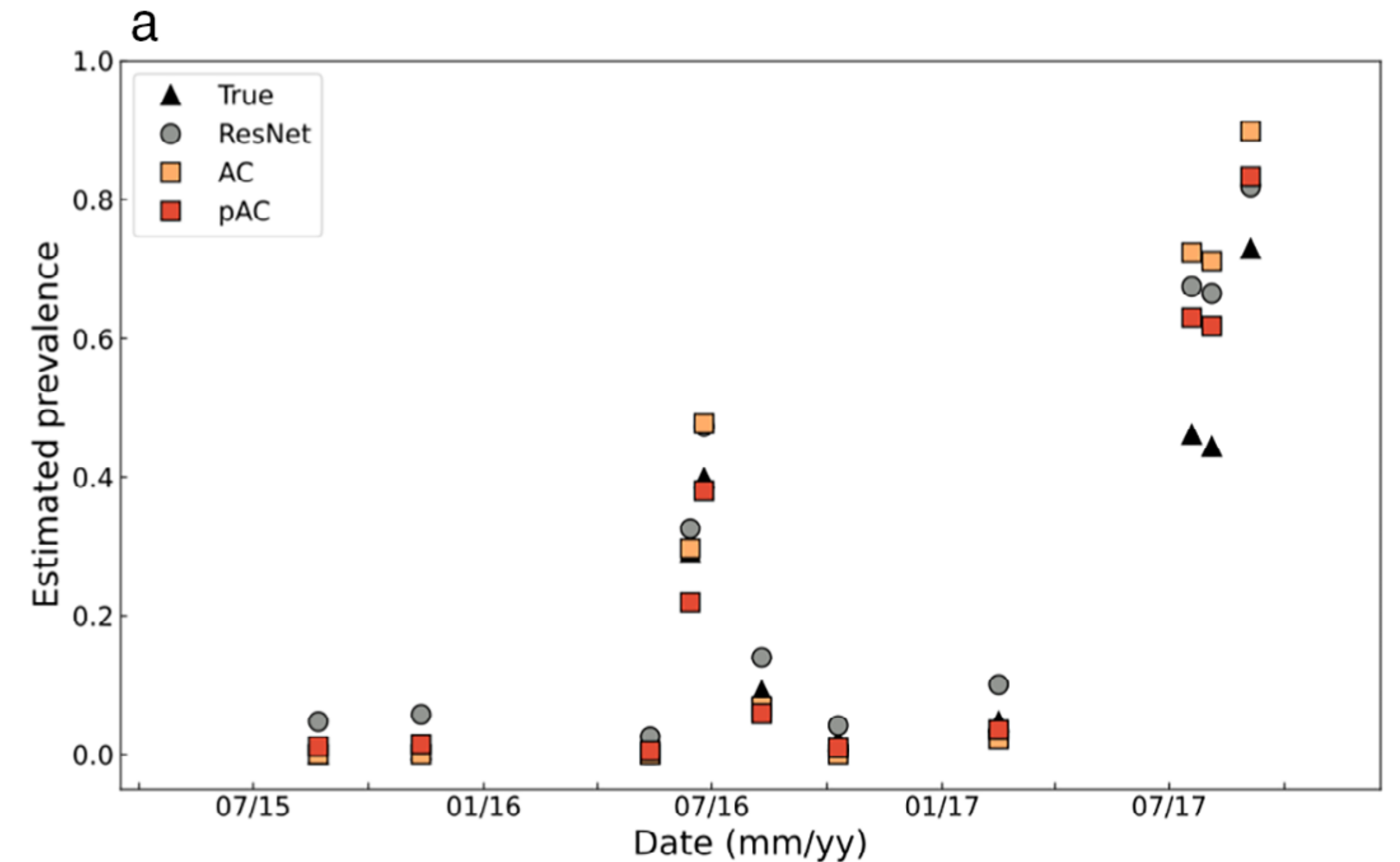
Limitations of machine learning applications in marine ecology

Feeling that machine/deep learning is **less explainable** than traditional statistics

Few dominant species/events, many rare ones

⇒ **Dataset shift**

⇒ **Quantification**



Limitations of machine learning applications in marine ecology

Feeling that machine/deep learning is **less explainable** than traditional statistics

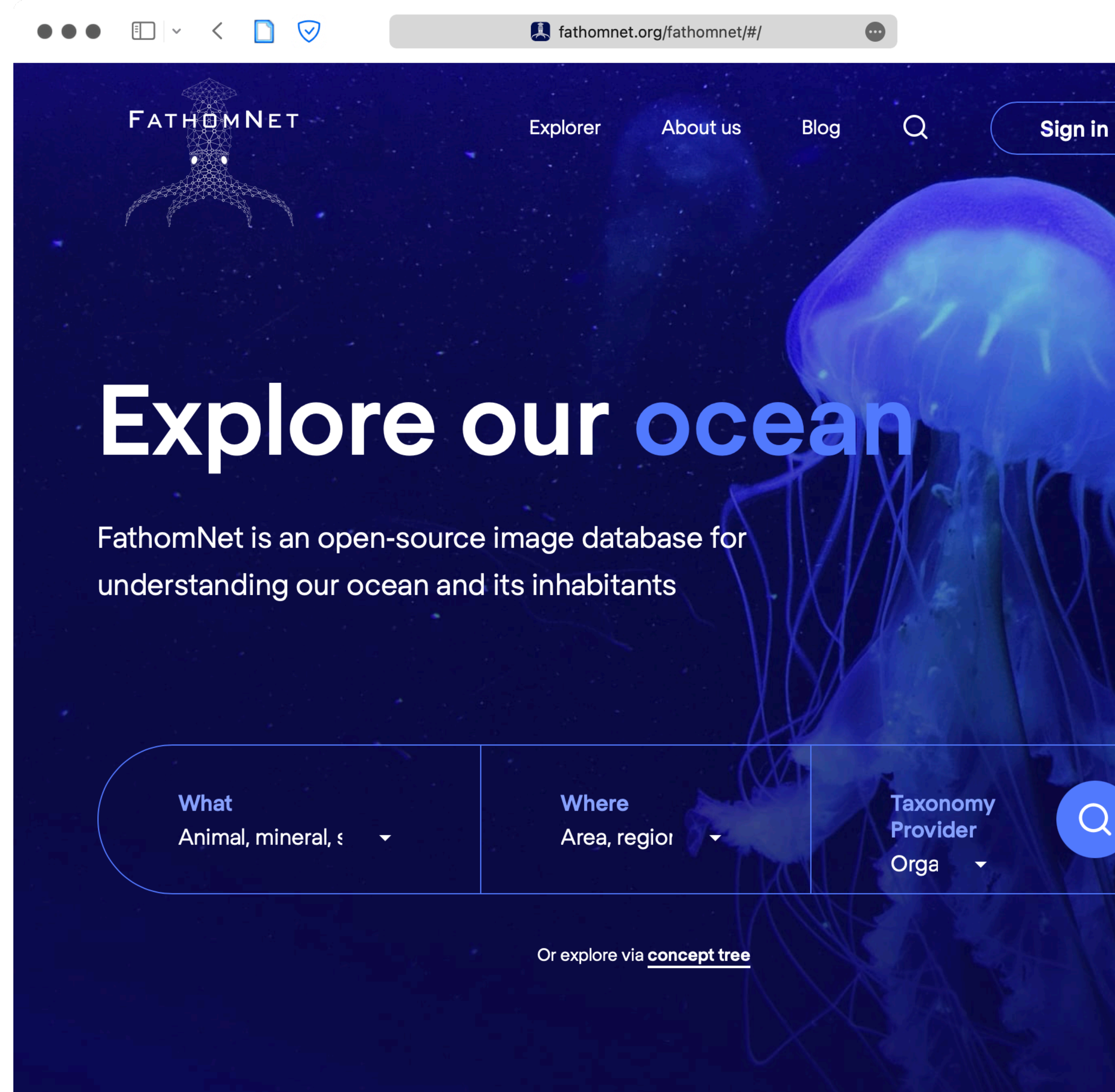
Few dominant species/events, many rare ones

⇒ **Dataset shift**

⇒ **Quantification**

Quality of **training sets**

⇒ Need more **sharing** of quality controlled datasets



Together we can

What's next?

New techniques

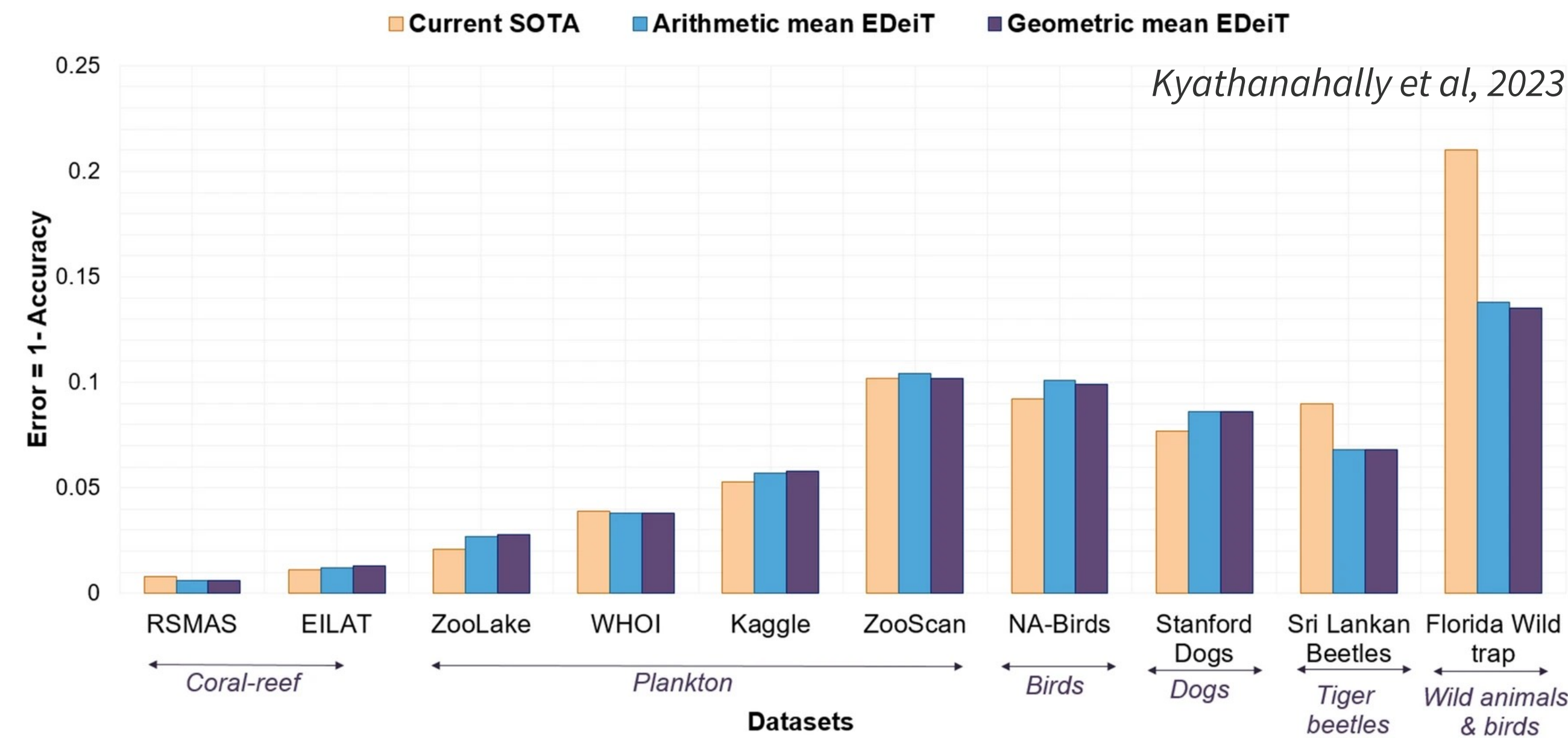
CNNs → Vision Transformers

GANs → Diffusion models

Newer models are often more **expensive** (in data, in compute) ⇒ are they worth the trouble?

Unlocking ML's potential requires cross field **collaboration**

⇒ We must train a **new generation** of scientists!



Going further

Join the group to update the database!



<https://www.zotero.org/groups/2325748/wgmlearn/>

irisson@normalesup.org



ICES
CIEM

BELMONT
FORUM

Shameful plug

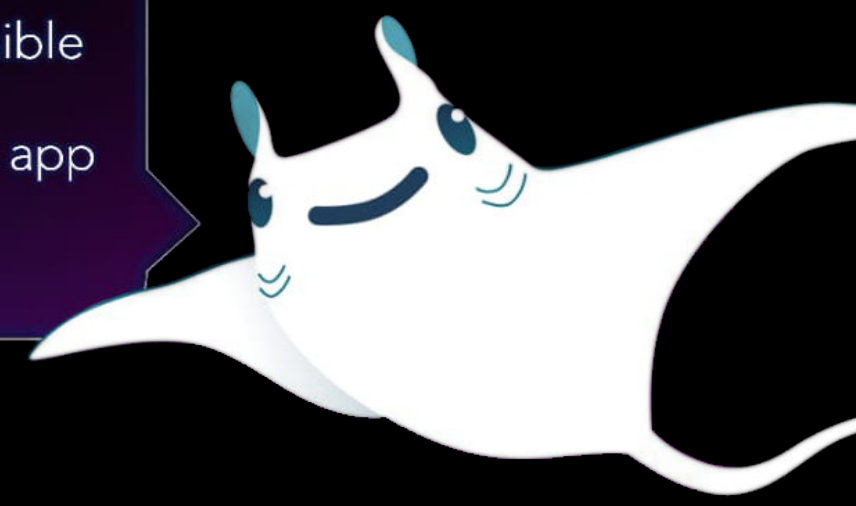
Aidez mon équipe à étudier les effets du changement climatique sur les écosystèmes marins en jouant à



le tout premier jeu de **RESOLV'IT**



Disponible
sur vos app
stores

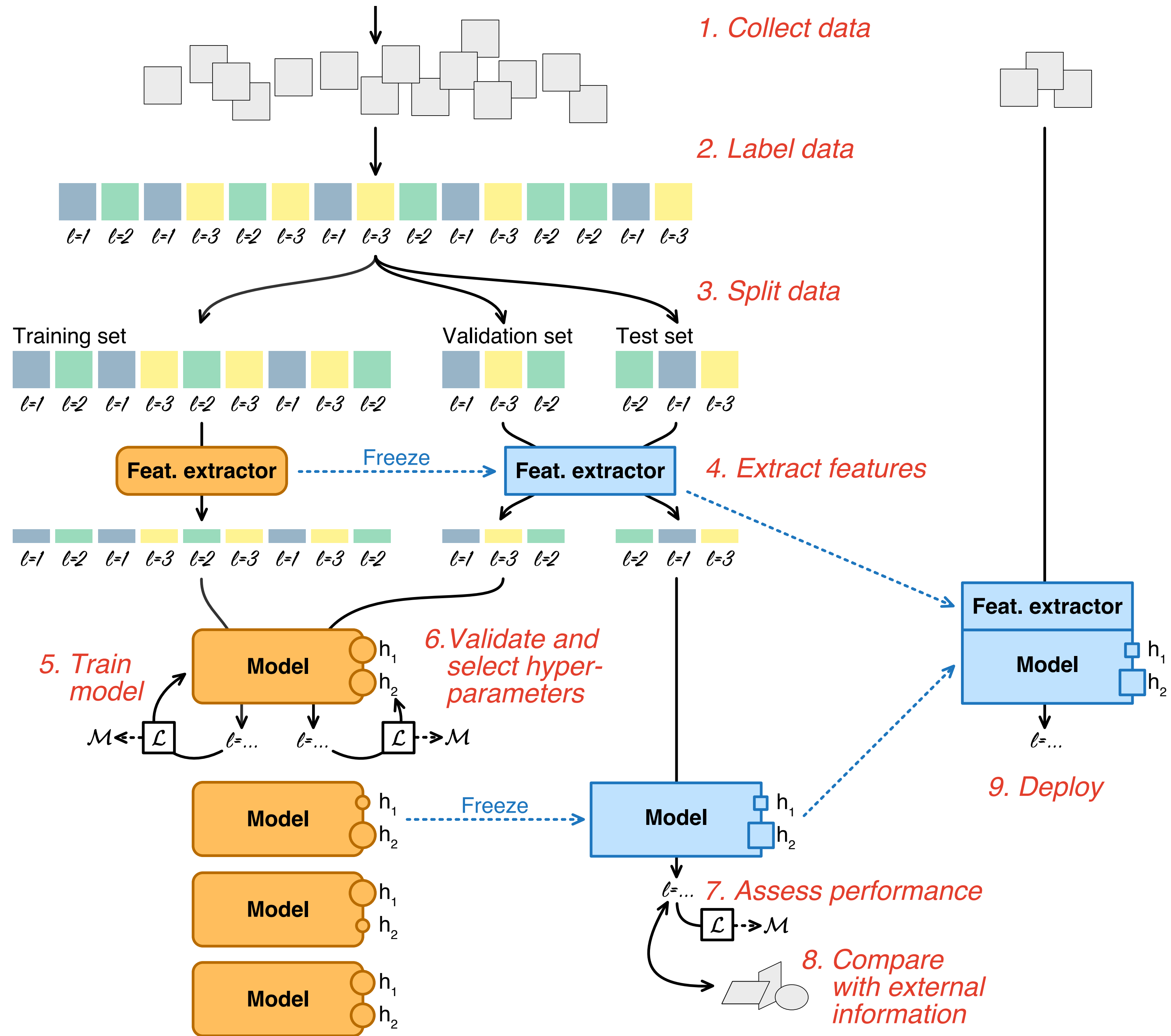


A ~~The~~ "definition" of machine learning

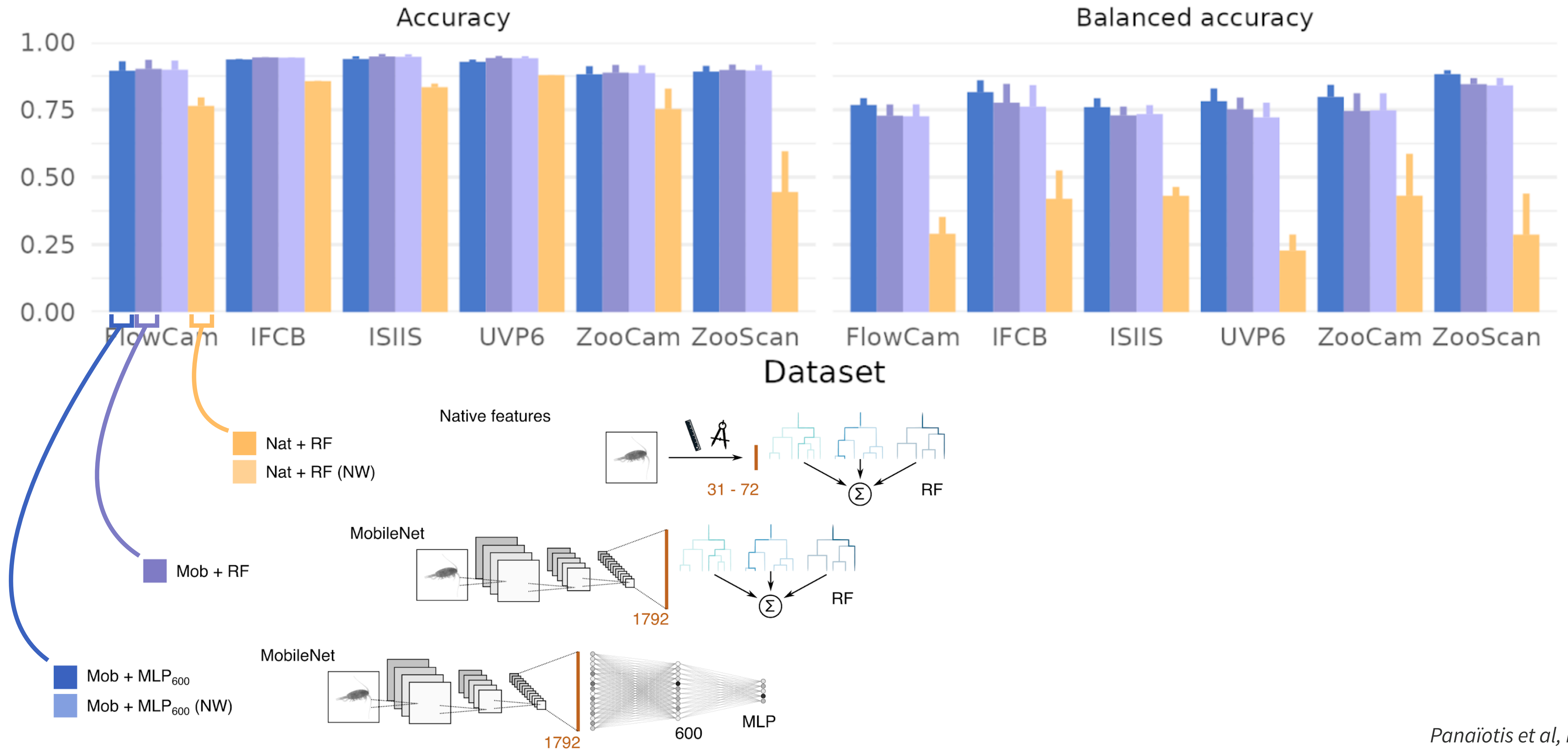
Has a **training** and **test** set

Feature extraction, data transformation, model fitting, and parameters tuning are done on the **training set only**

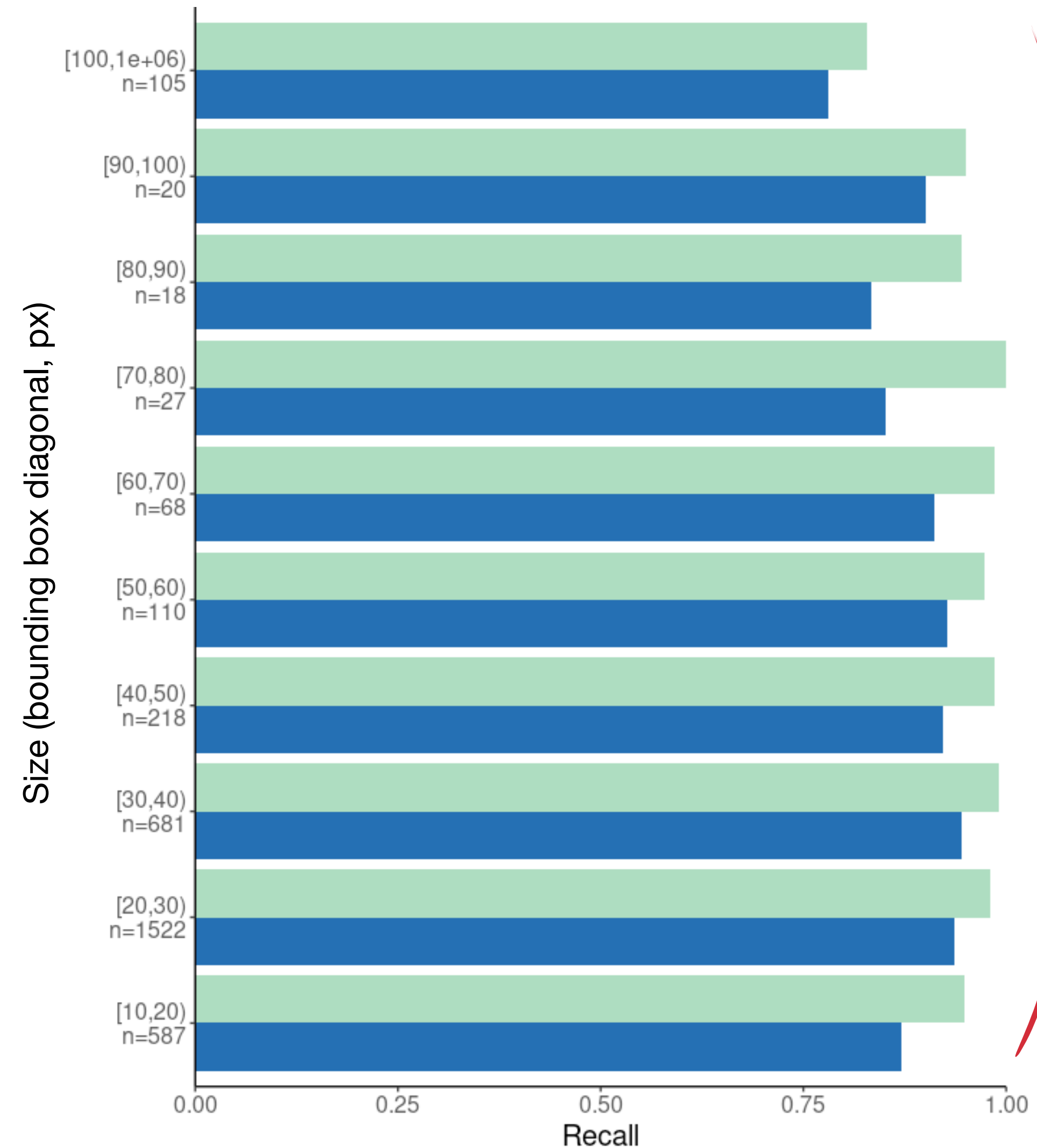
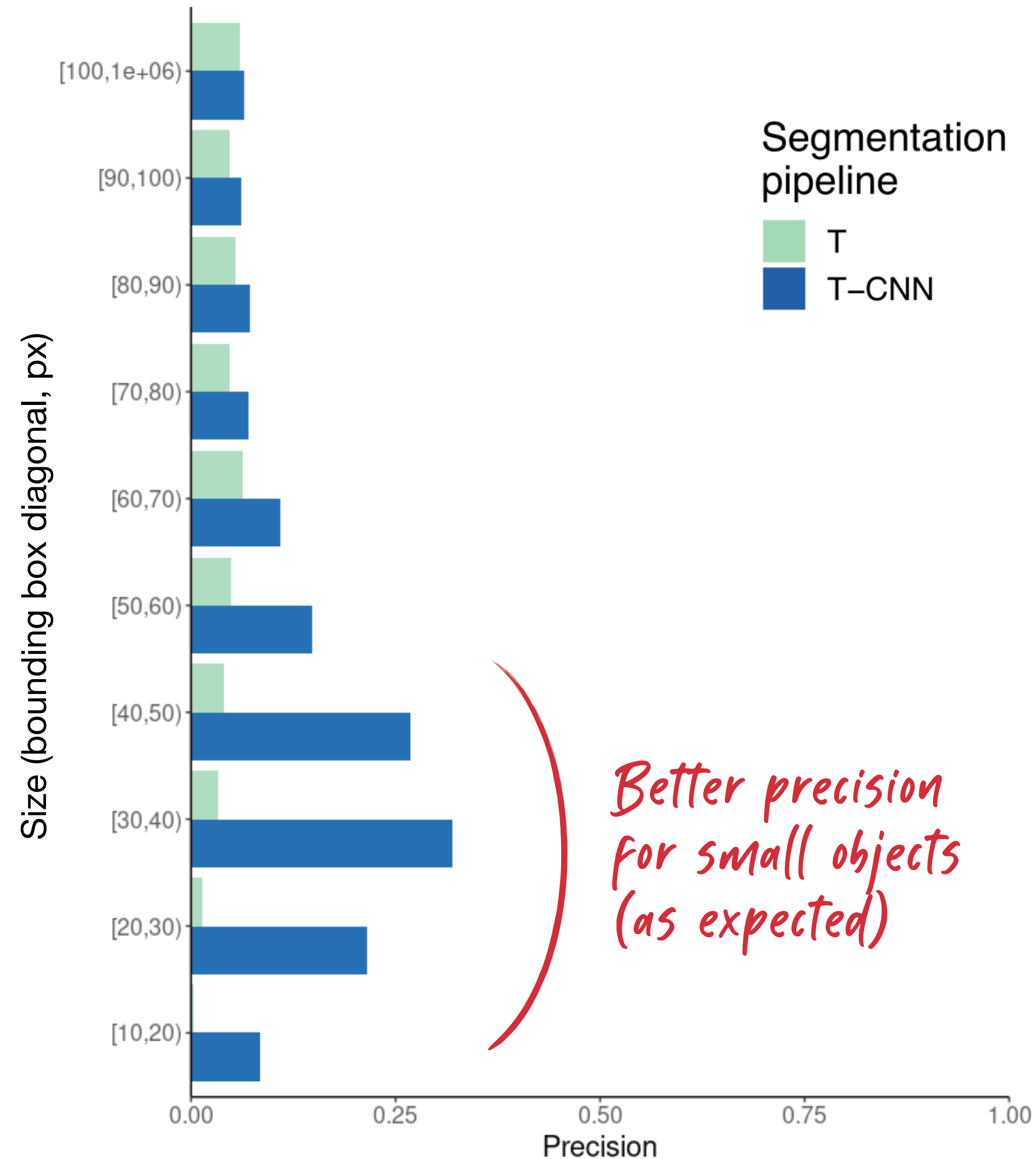
The model is then **applied** to the test set



Why? It's the features, not the classifier

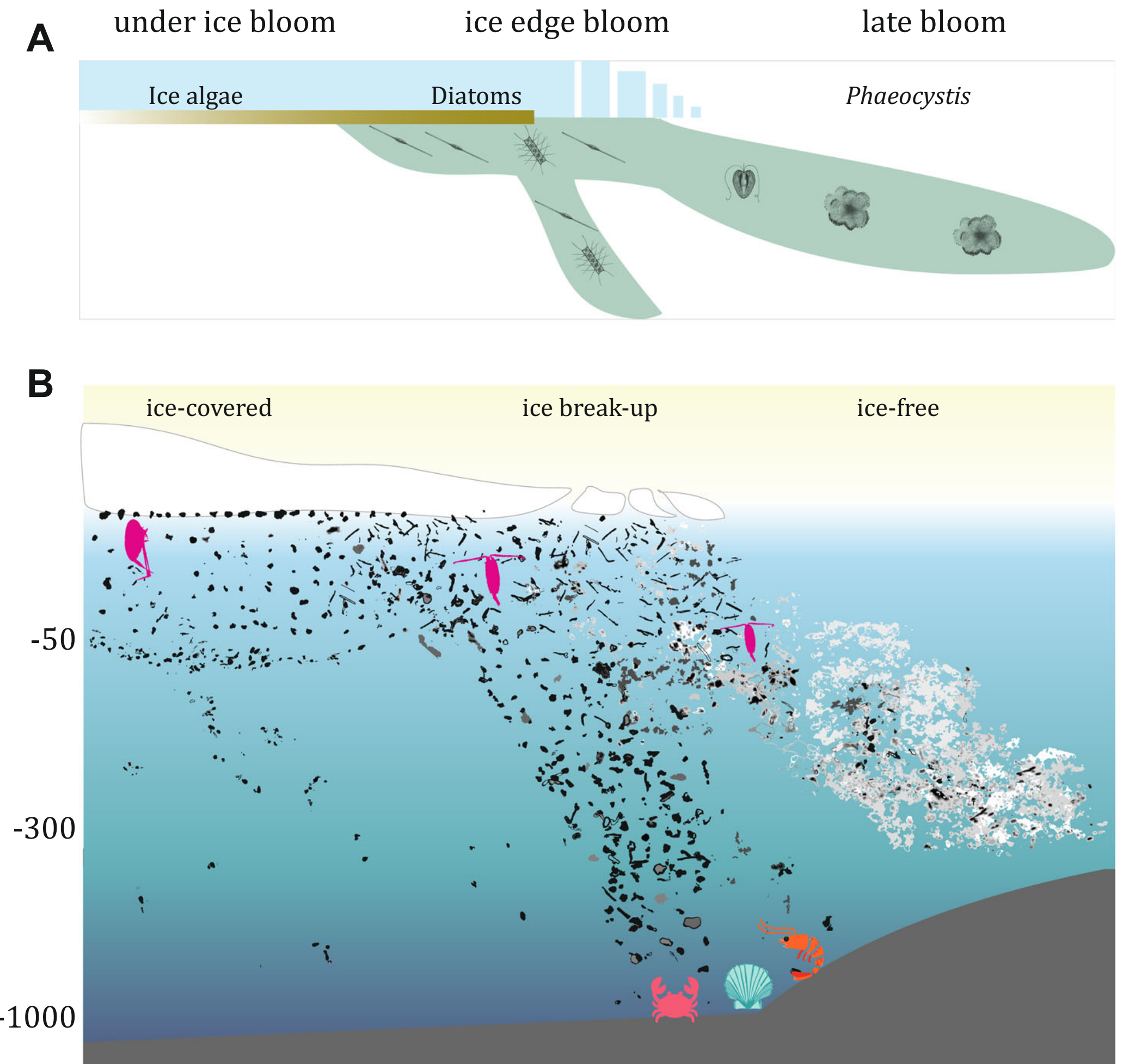
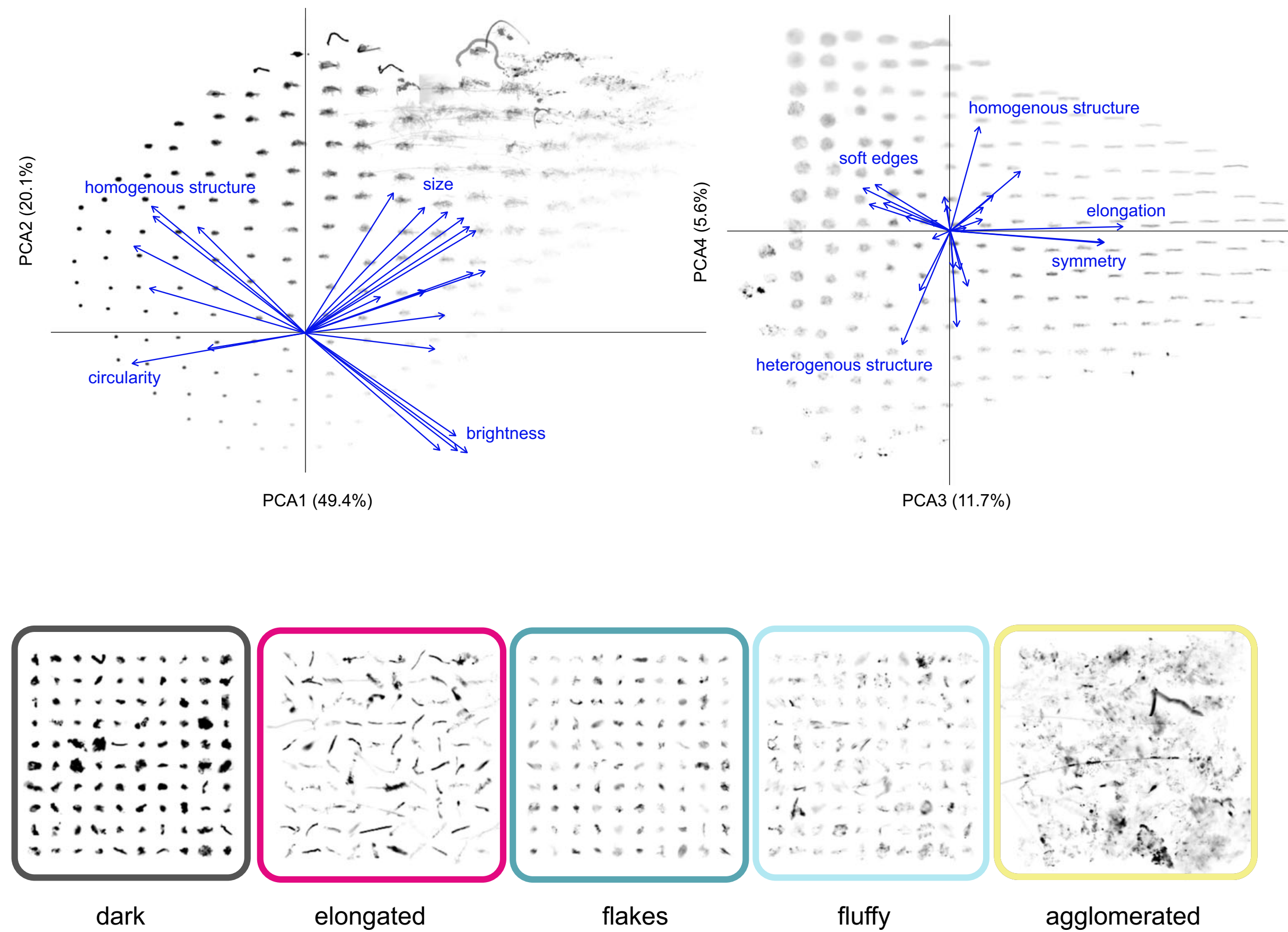


... with little influence of size on performance

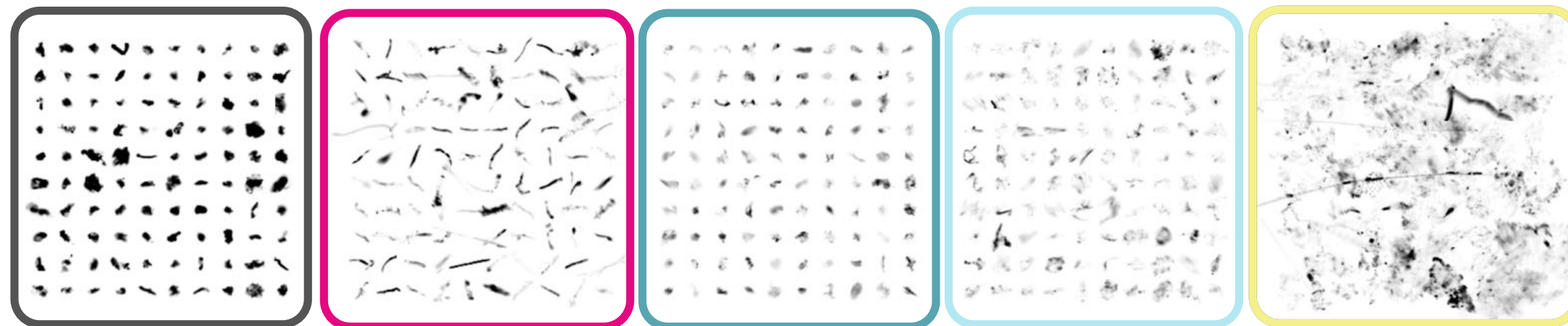
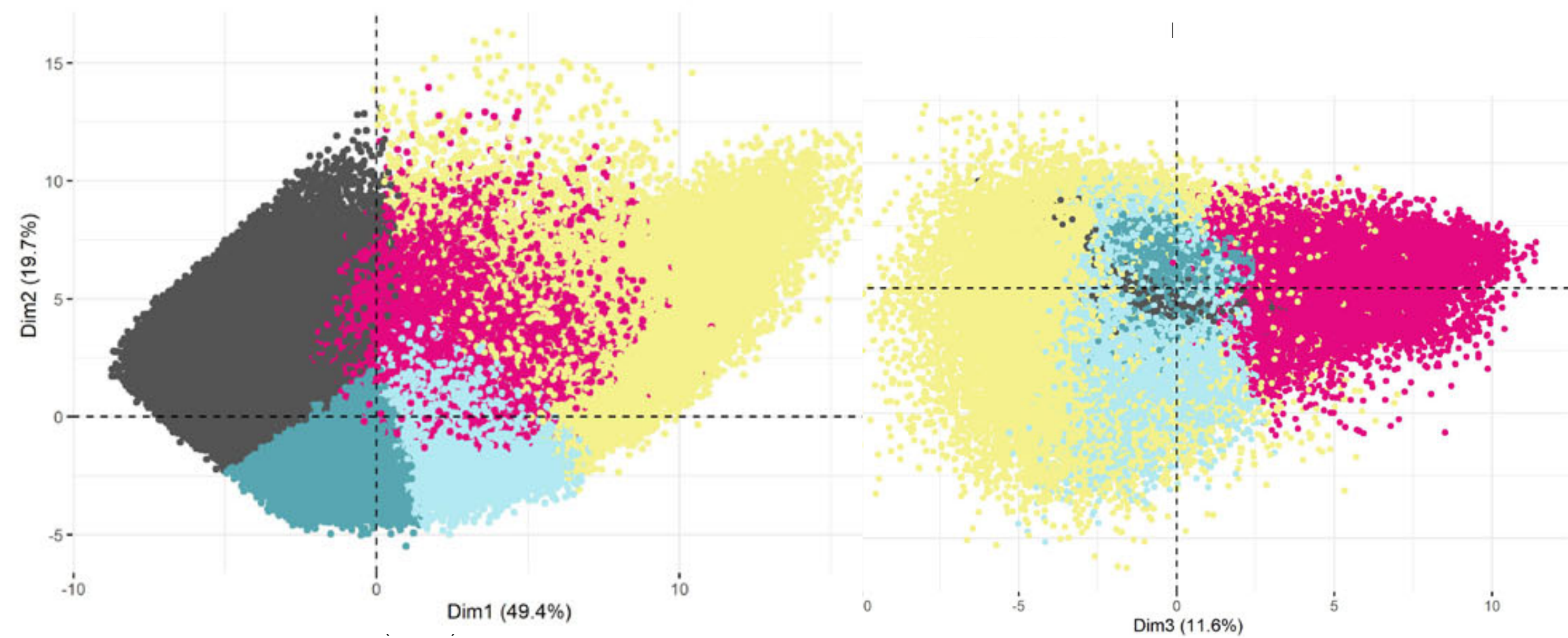


*Little effect
of size on
recall*

Unsupervised clustering in a morphological space



Unsupervised clustering in a morphological space



dark

elongated

flakes

fluffy

agglomerated

