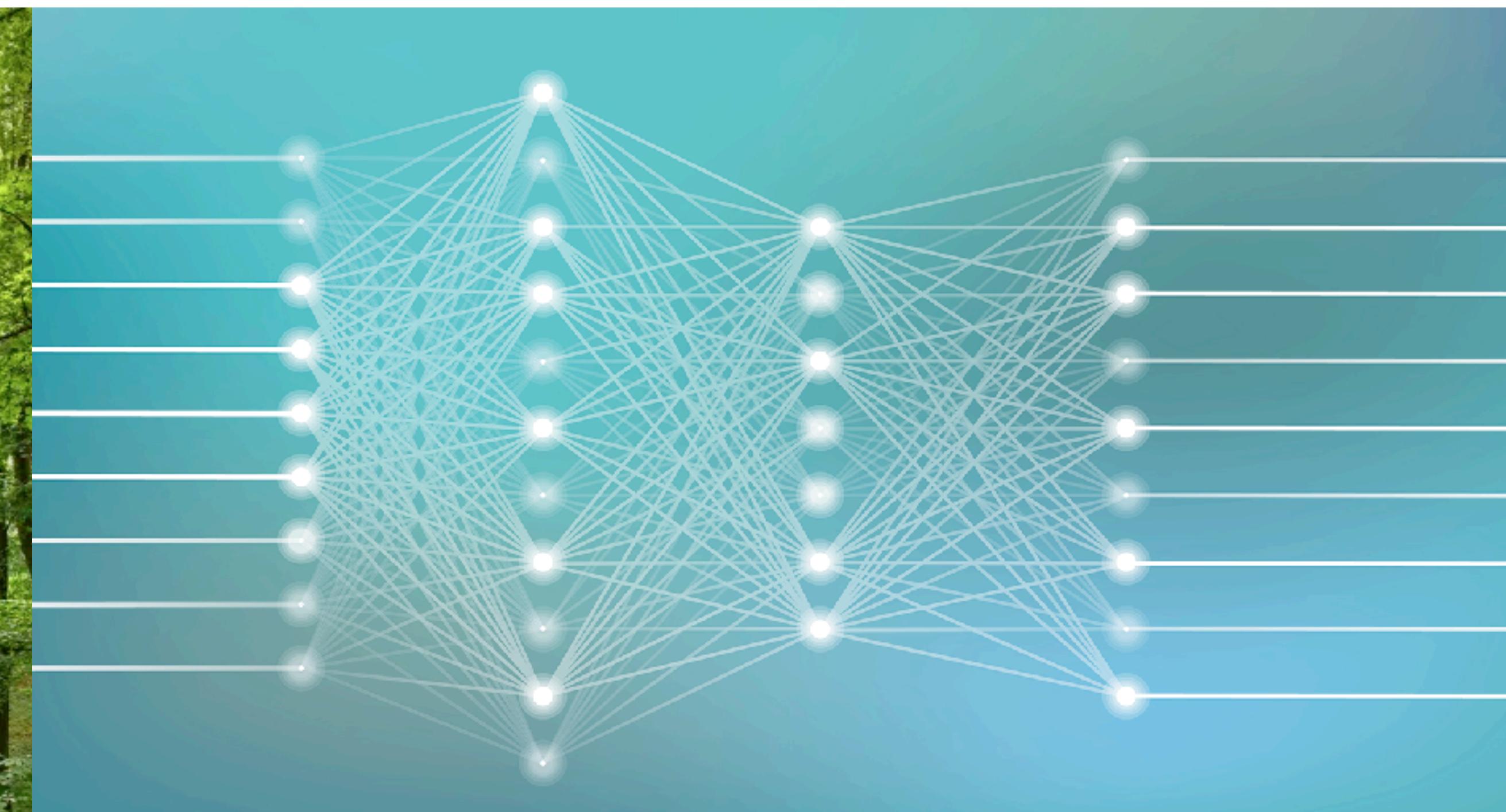
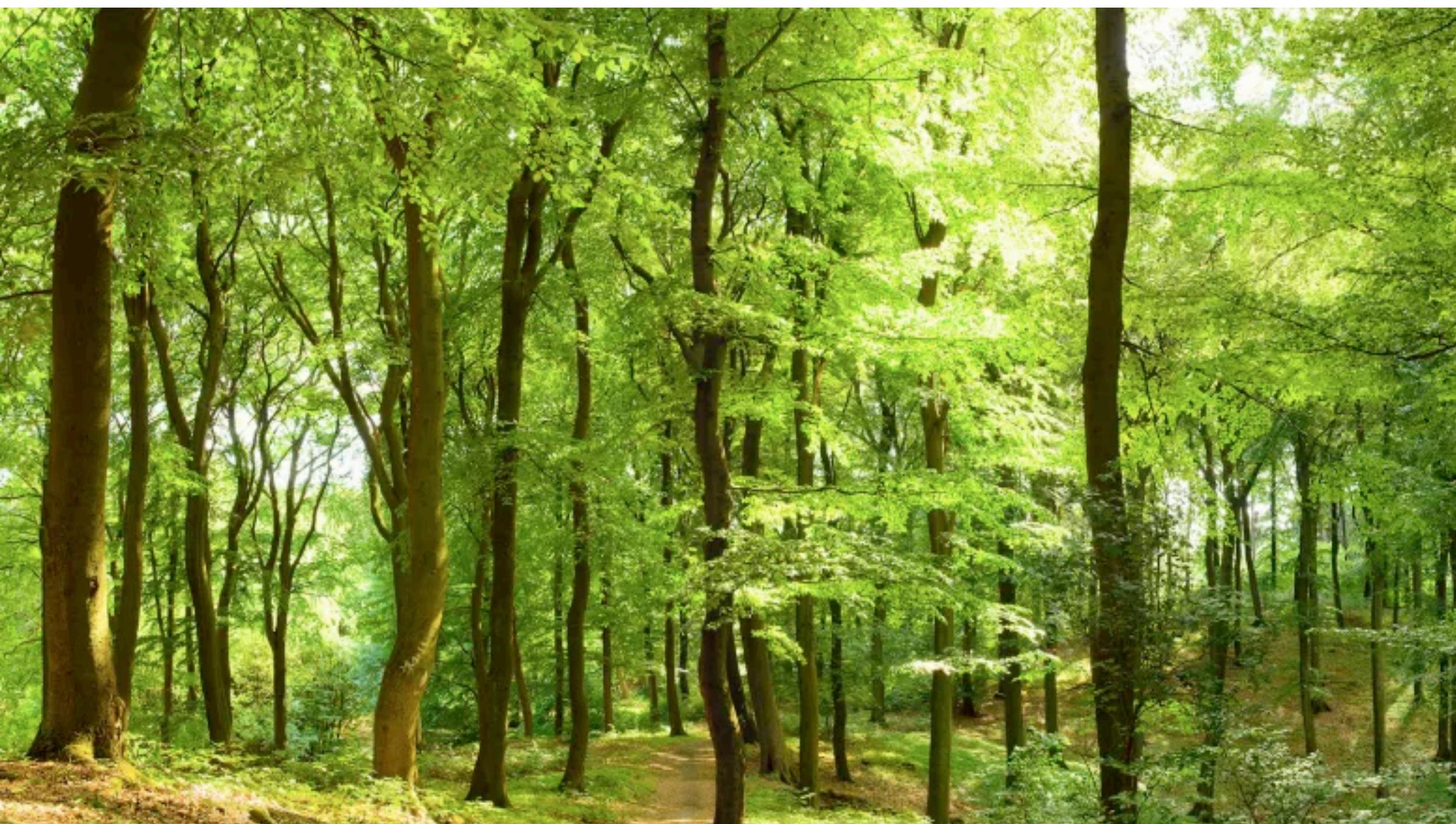


Benchmark of plankton image classification: CNN improve detection of rare taxa

T Panaïotis, G Boniface-Chang, G Dulac-Arnold, T Biard, B Blanc, L Caray--Counil, C Desnos, A Elineau, L Jalabert, R Kiko, F Lombard, M Picheral, JB Romagnan, L Stemmann, JO Irisson

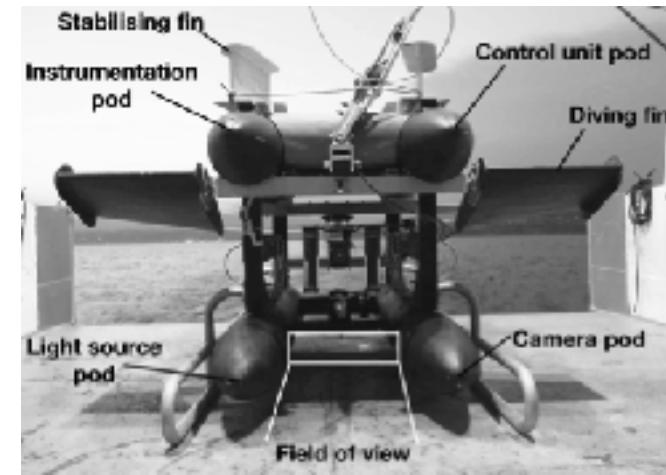
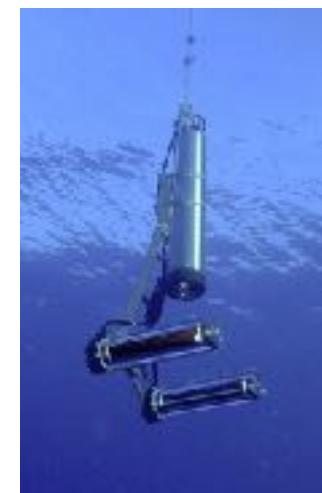
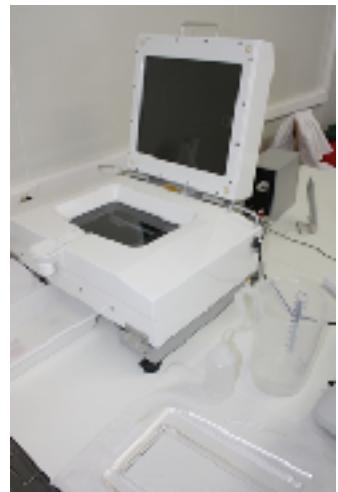
ASLO 2021 Aquatic Sciences Meeting



Context

Plankton imaging

More instruments



More data

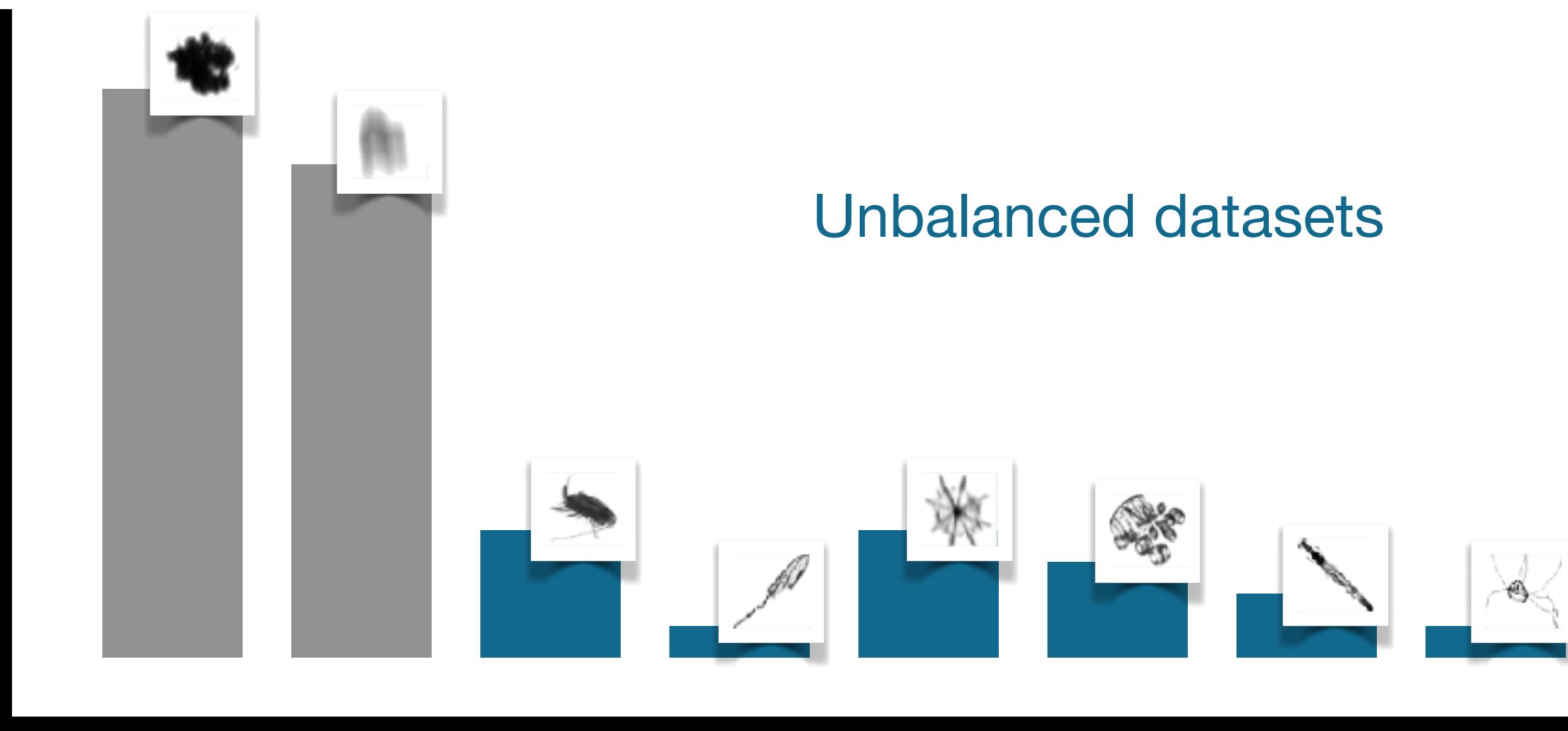


Data processing bottleneck



- Report performances of two classification approaches
- Provide baseline results for future comparisons

Unbalanced datasets



174 papers



3 benchmark datasets

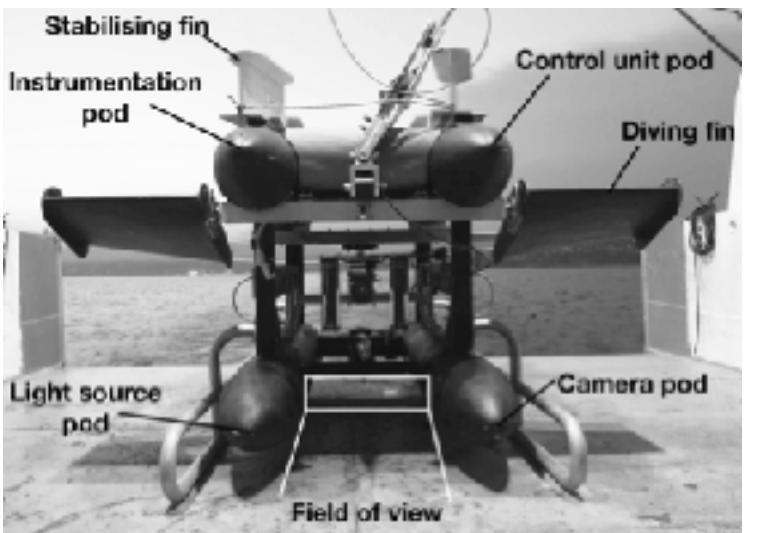


Lack of inter-comparison

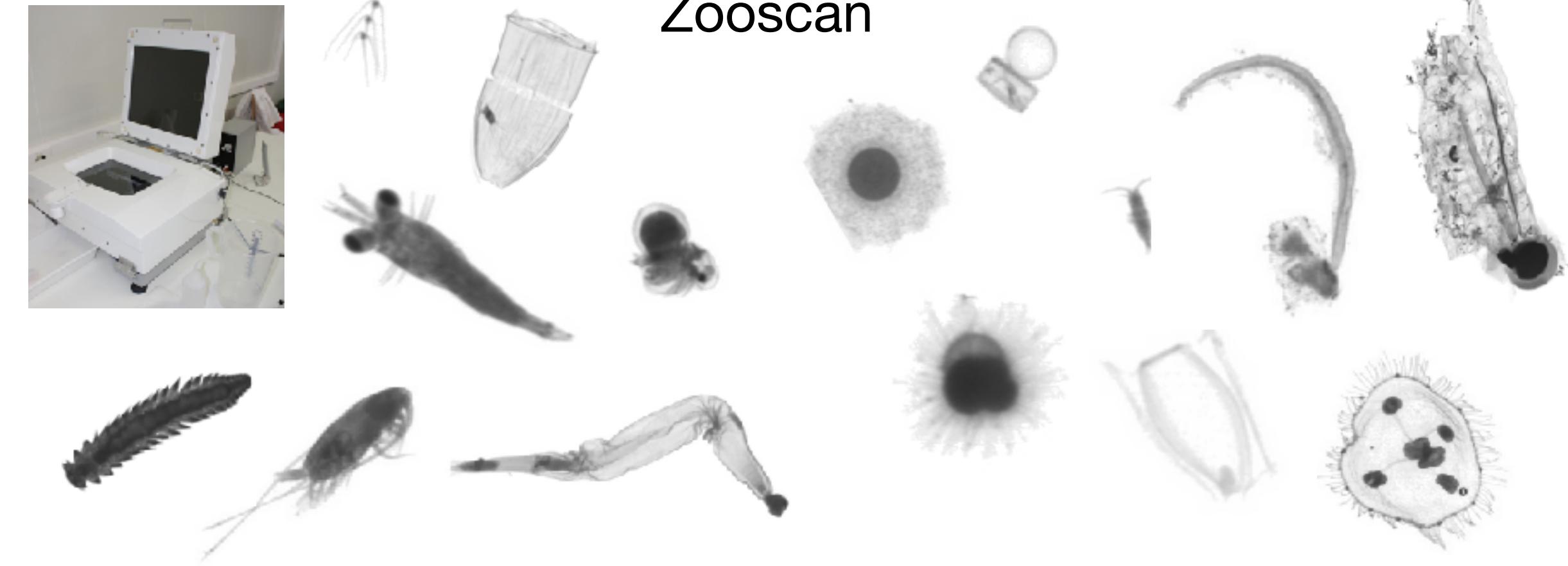
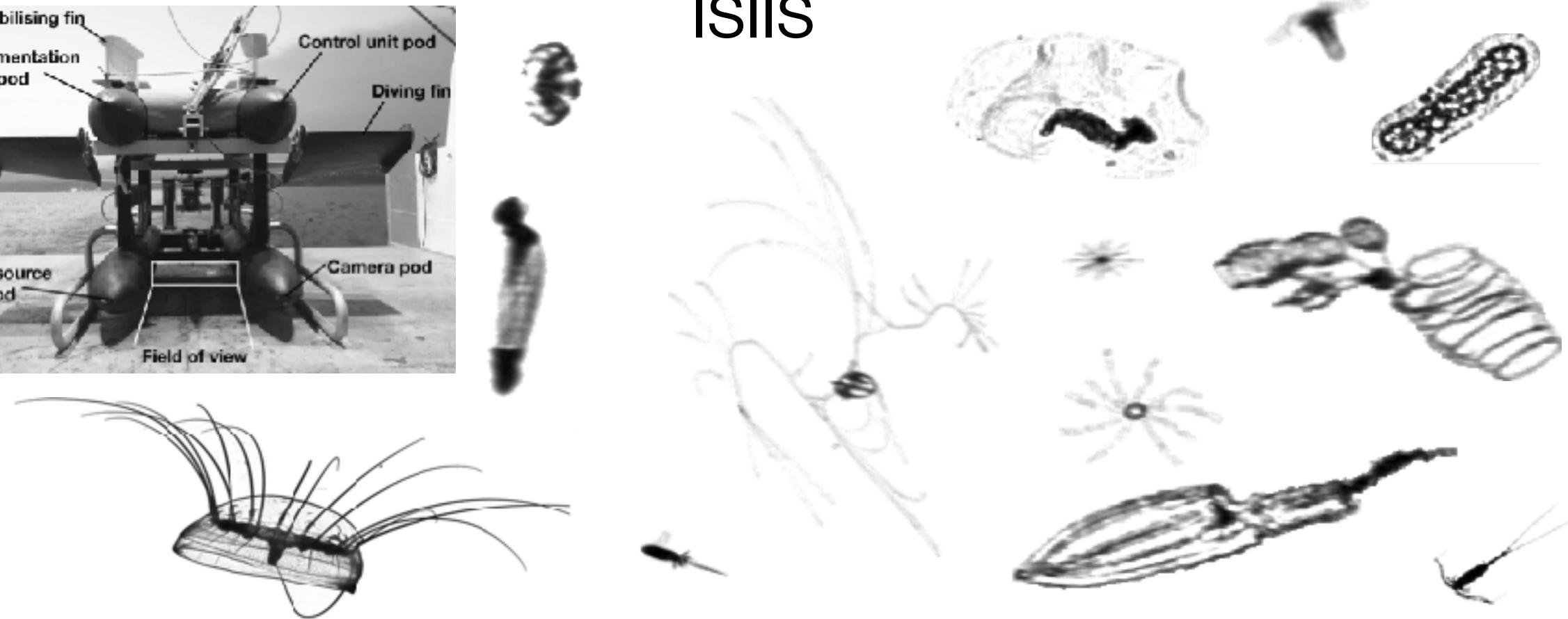


Material & methods

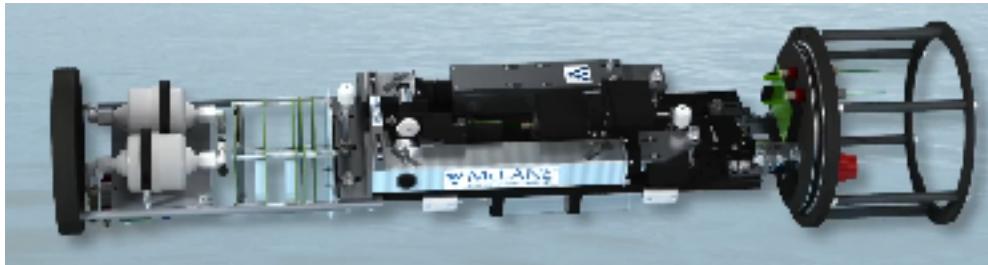
Datasets – imaging instruments



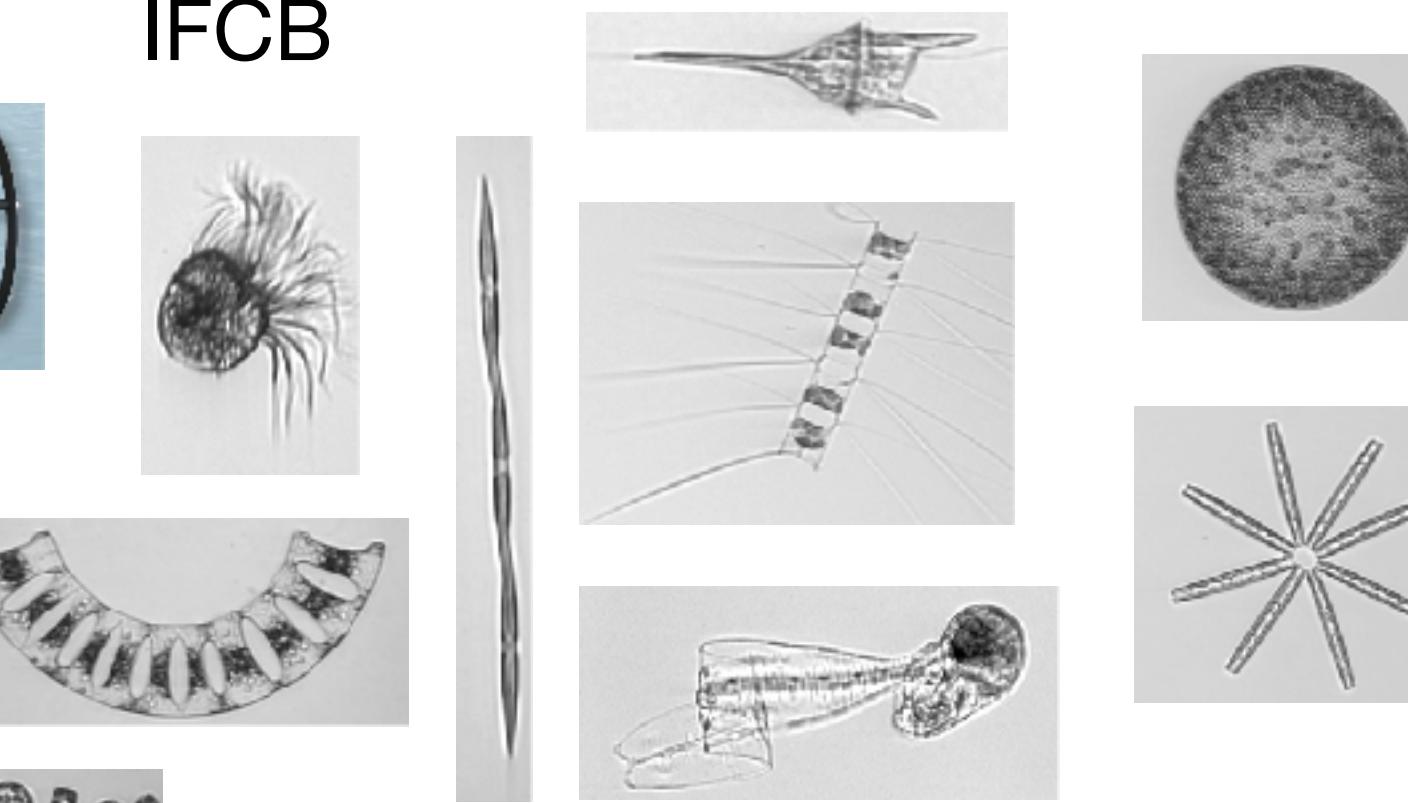
ISIIS



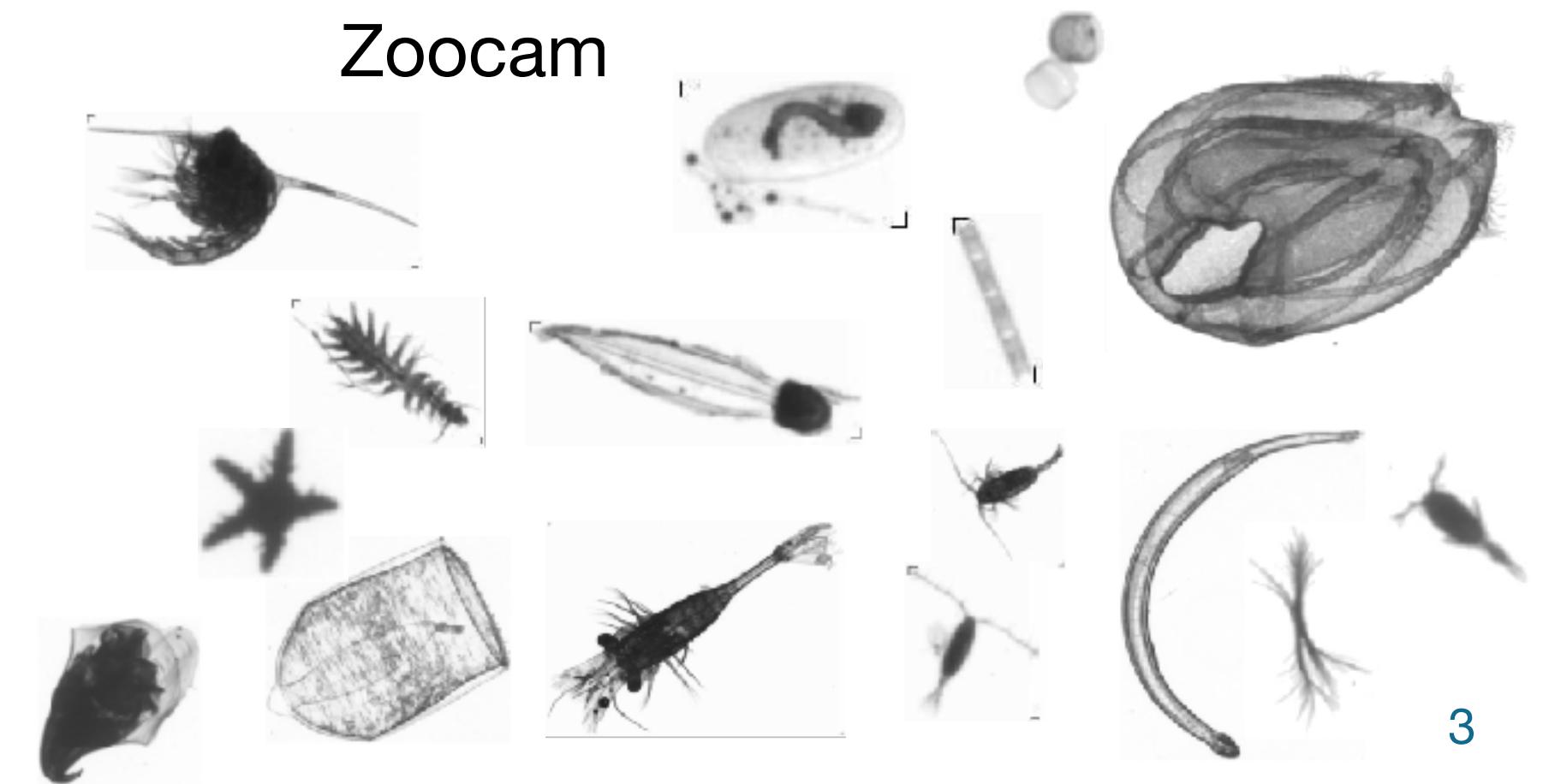
Zooscan



IFCB

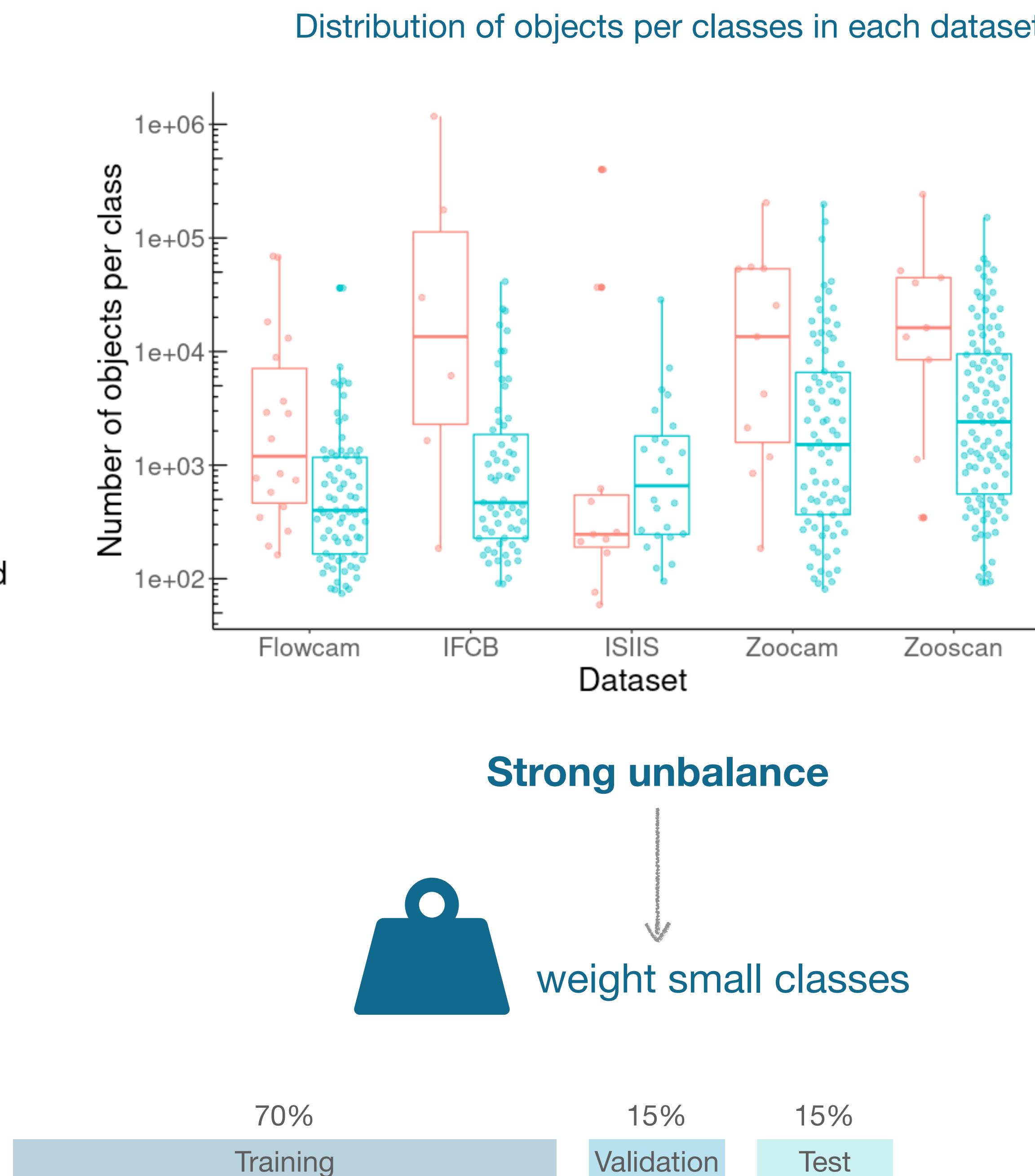
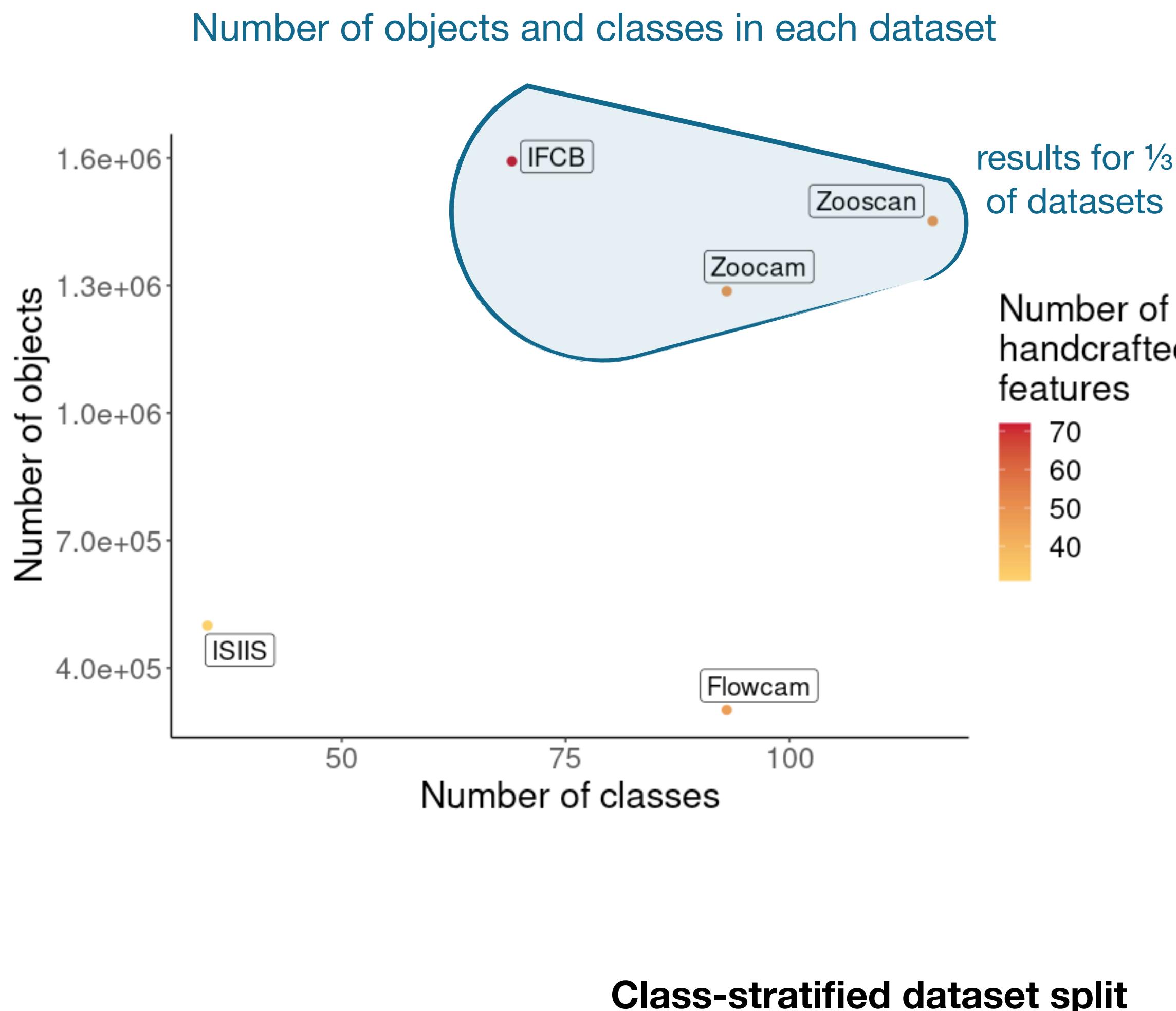


Zoocam



Material & methods

Datasets – composition



Material & methods

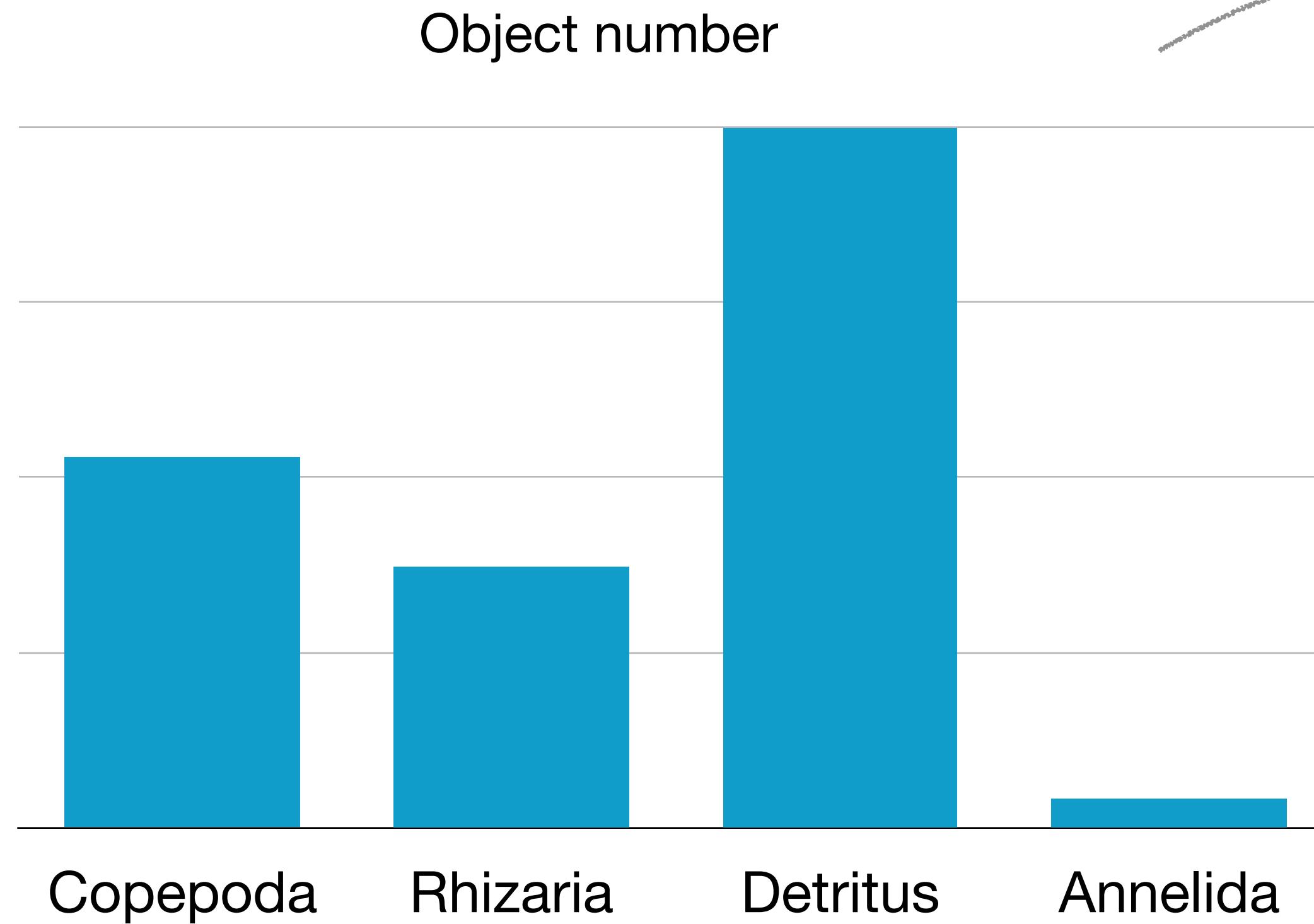
Datasets – processing

- include **every** object
- as **detailed as possible**... but small classes
 - ▶ if possible, merge with close relevant class
 - ▶ otherwise, merge with miscellaneous
- define **broader classes** corresponding to traditional approaches (e.g. Copepoda)
- categorise classes in **plankton classes** or not

levelName	level0	nb_level0	level1	nb_level1
-Pontellidae	Pontellidae	239	Pontellidae	239
-Rhincalanidae	Rhincalanidae	1	Calanoida	197751
^-Temoridae	Temoridae	14638	Temoridae	14638
-copepoda-cut	copepoda-cut	9686	Copepoda	18557
-Cyclopoida			Cyclopoida	139175
^-Oithonidae	Oithonidae	129334	Cyclopoida	139175
-oithonidae-badfocus	oithonidae-bafocus	4111	Cyclopoida	139175
-oithonidae-face-view	oithonidae-face-view	4569	Cyclopoida	139175
^-oithonidae-side-view-no-ante	oithonidae-side-view-nc-antena	1181	Cyclopoida	139175
-dead	dead<Copapoda	17243	empty_Copepoda	17243
-Harpacticoida	Harpacticoida	645	Harpacticoida	645
-dead	dead<Harpacticoida	619	empty_Harpacticoida	619
-Euterpnina	Euterpnina	1449	Euterpnina	1449
^-Microsetella	Microsetella	318	Microsetella	318
-multiple	multiple<Copepoda	5750	multiple_Copepoda	5750
^-Poecilostomatoida	Poecilostomatoida	679	Poecilostomatoida	692
-Corycaeidiae	Corycaeidiae	3859	Corycaeidiae	3859
-Oncaeidae				
^-Oncea	Oncea	23336	Oncea	23336
^-Sapphirinidae	Sapphirinidae	13	Poecilostomatoida	692
^-Thecostraca				
^-Cirripedia				
-cypris	cypris	882	Cirripedia	8287
^-nauplii	nauplii<Cirripedia	7425	Cirripedia	8287
-nauplii	nauplii<Crustacea	4637	nauplii_Crustacea	4637
-Oligostraca				
^-Ostracoda	Ostracoda	73	other_living	25454
^-tail	tail<Crustacea	587	part_Crustacea	1264
^-Hexapoda				
^-Insecta	Insecta	2	detritus	204132
-Bryozoa				
^-cyphonaute	cyphonaute	1593	cyphonaute	1593
-Chaetognatha	Chaetognatha	3086	Chaetognatha	3150
^-tail	tail<Chaetognatha	54	Chaetognatha	3150
^-Chordata				
-Cephalochordata				
^-Branchiostomidae				
-Branchiostoma	Branchiostoma	18	other_living	25454
-Cranata				
^-Vertebrae				
-Gnathostomata				
^-Actinopterygii	Actinopterygi	100	Actinopterygii	100
-Clupeiformes temp				
^-Clupeidae temp				
^-Sardina temp				
^-egg unkn temp	egg unkn temp<Sardina temp	7754	egg_Sardina	7754
^-Engraulidae temp				
^-egg unkn temp	egg unkn temp<Engraulidae ter	41455	egg_Engraulidae	41455
-egg	egg<Actinopterygii	5138	egg_Actinopterygii	5138

Material & methods

Models – class weights



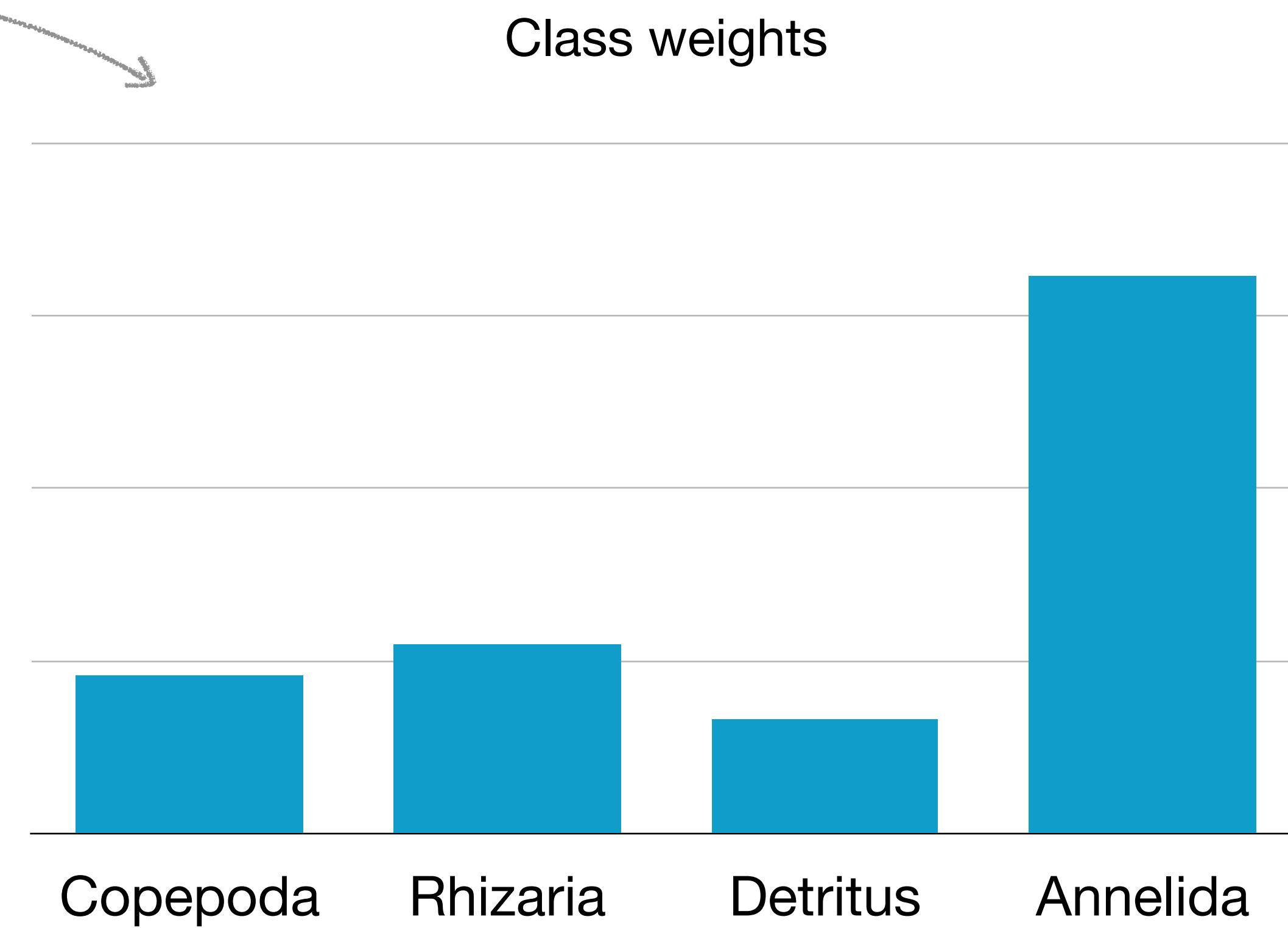
Weights for class i

$$W_i = \sqrt{\frac{c_{max}}{c_i}}$$

Number of objects in largest class

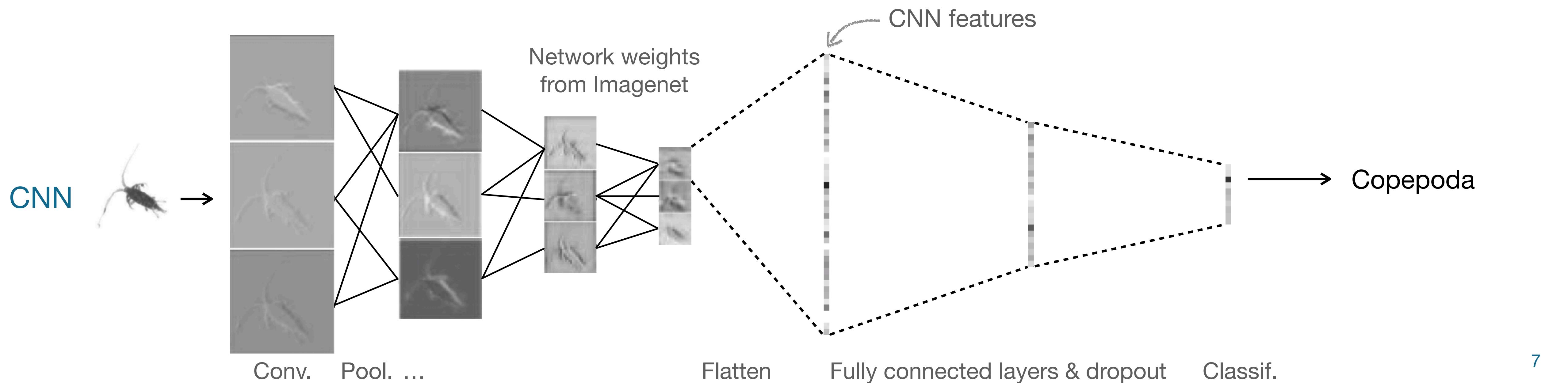
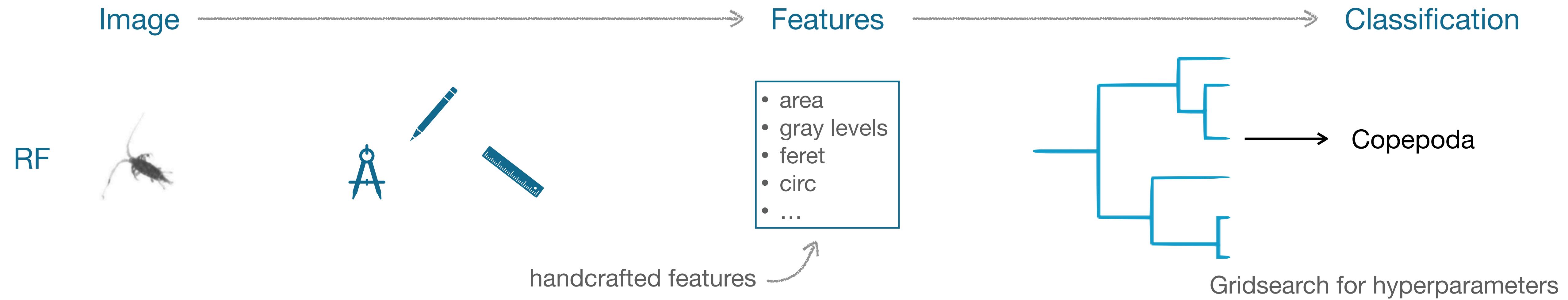
Number of objects in class i

The equation $W_i = \sqrt{\frac{c_{max}}{c_i}}$ represents the formula for calculating class weights. It consists of three parts: "Weights for class i " (the label above the equation), the equation itself, and "Number of objects in largest class" (the label pointing to the denominator). Below the equation is "Number of objects in class i " (the label pointing to the variable c_i).



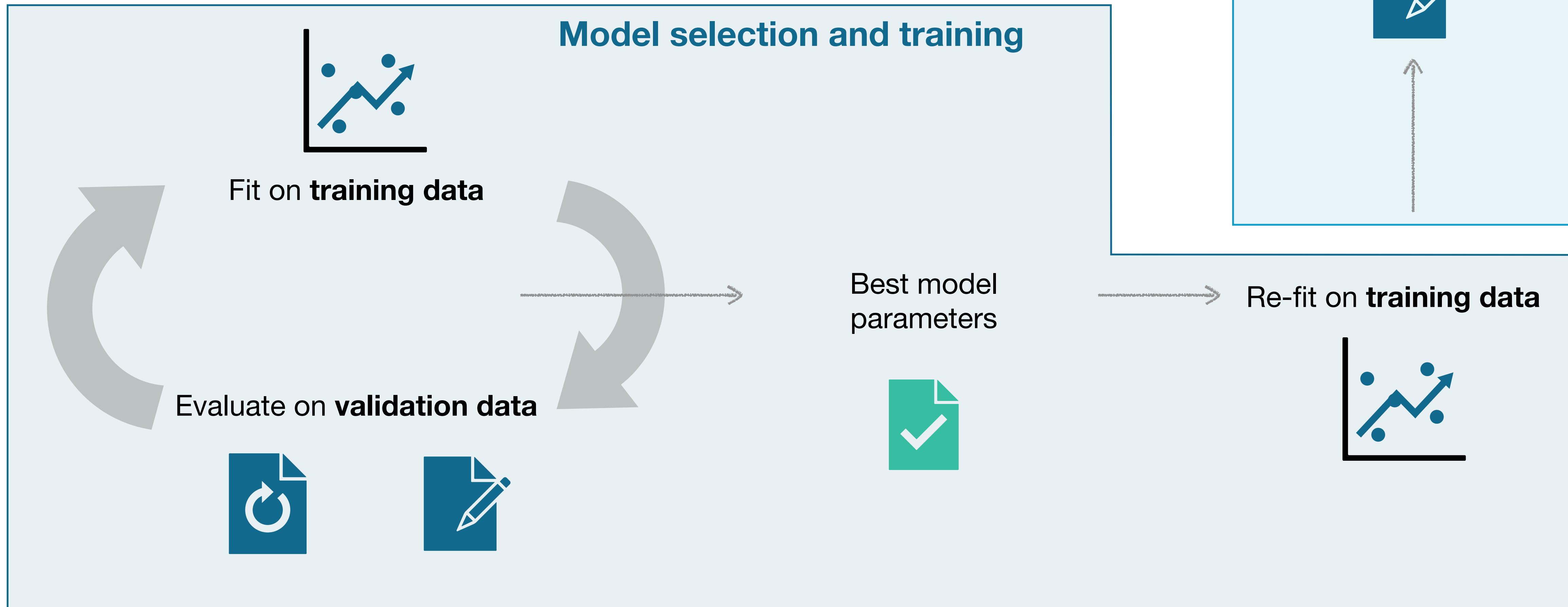
Material & methods

Models



Material & methods

Models – training procedure



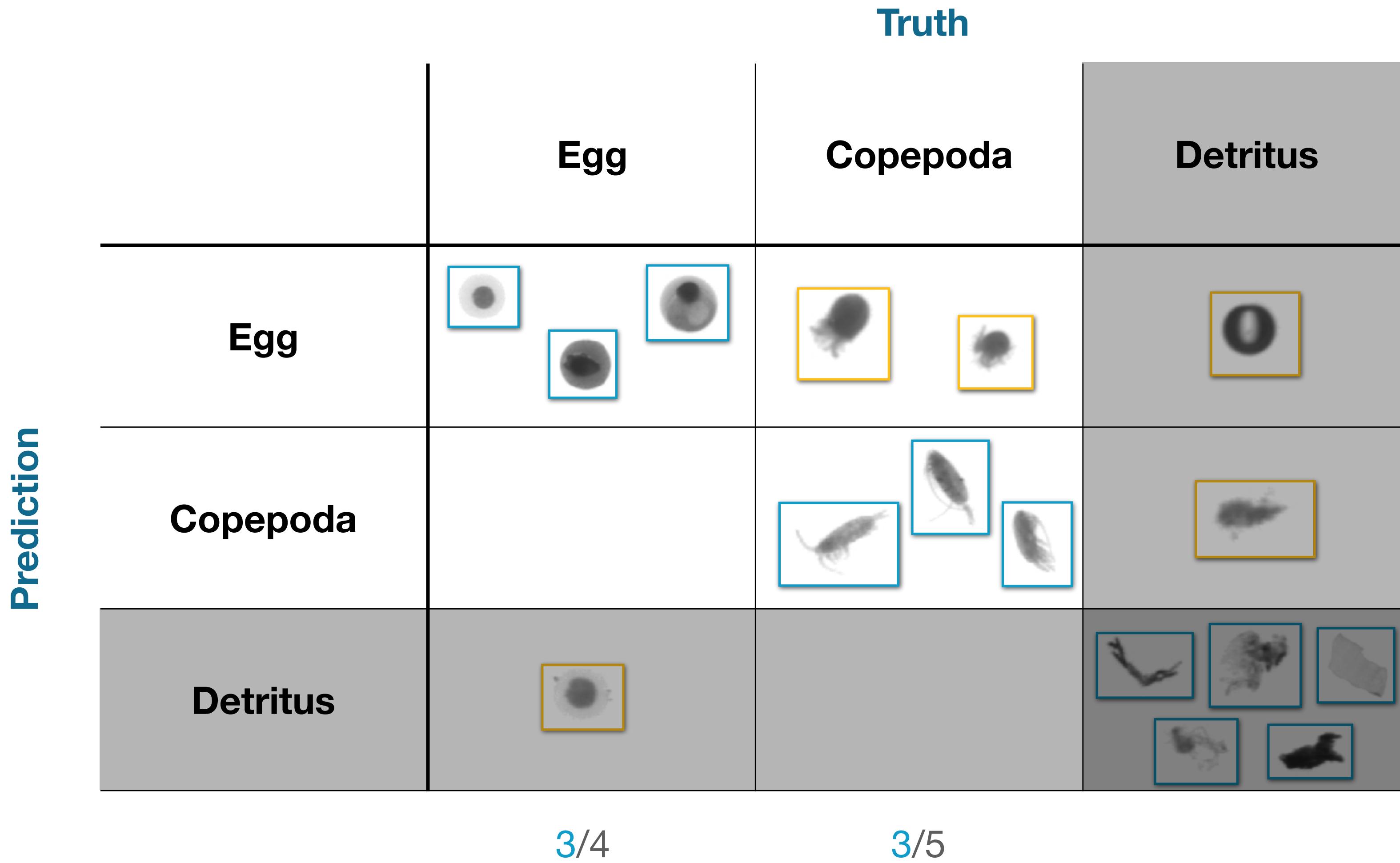
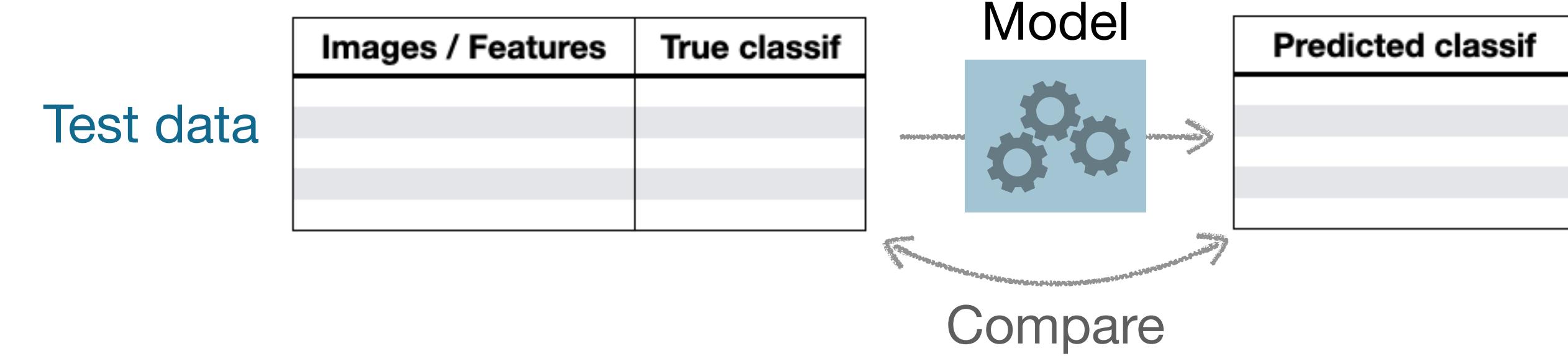
Model assessment

Evaluate on **test data**



Material & methods

Model evaluation



Accuracy

$\frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$

Balanced accuracy

Weighted
plankton
precision

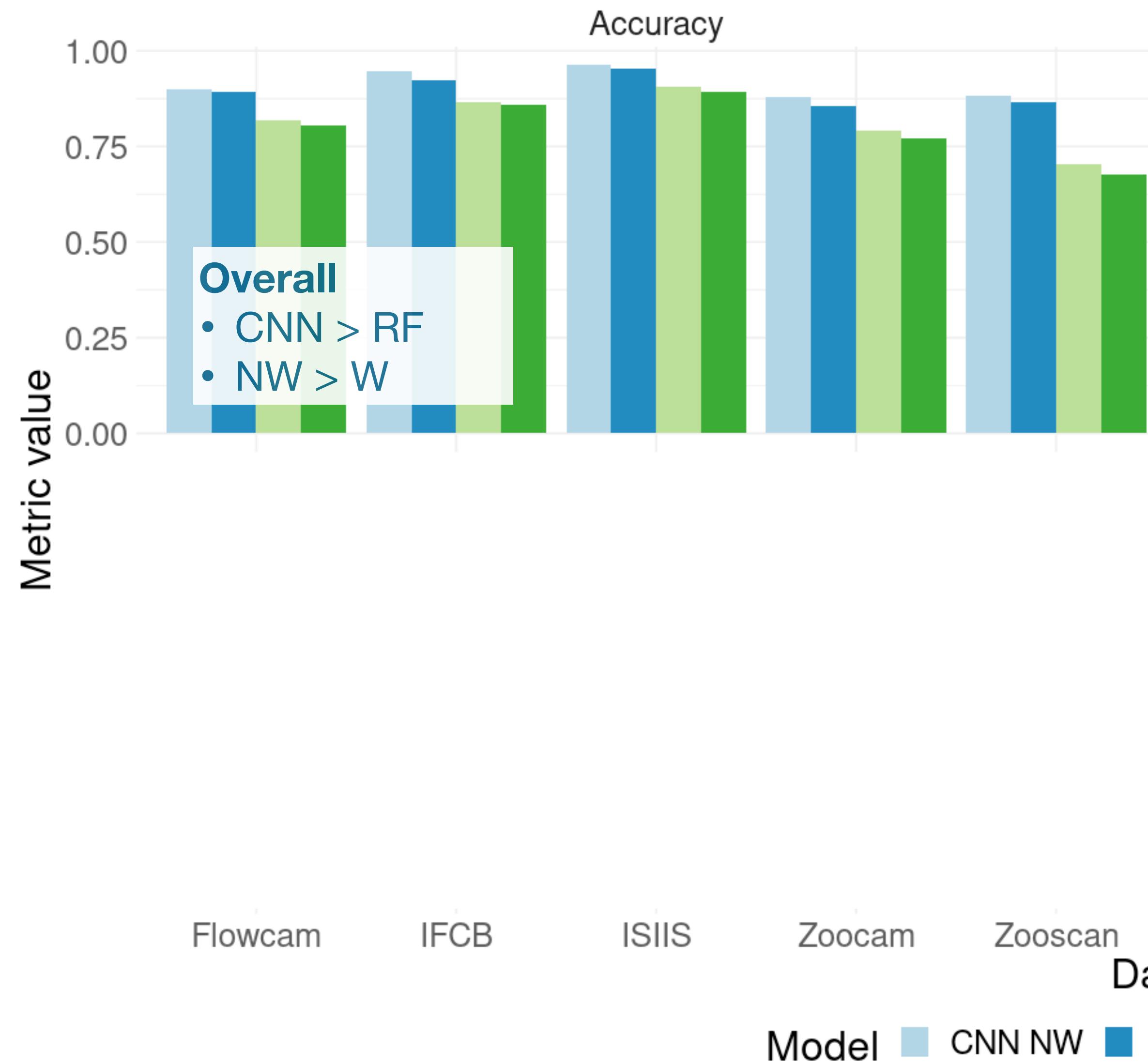
60%

Weighted
plankton
recall

67%

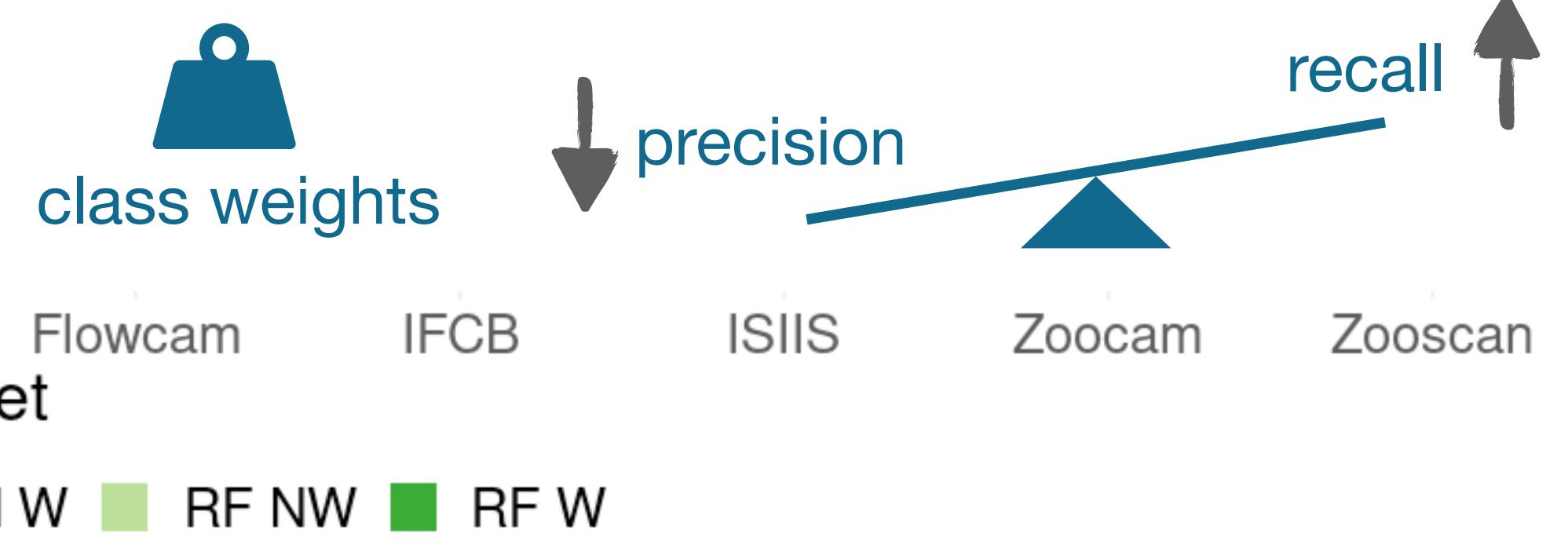
Results

Model performances on detailed classes



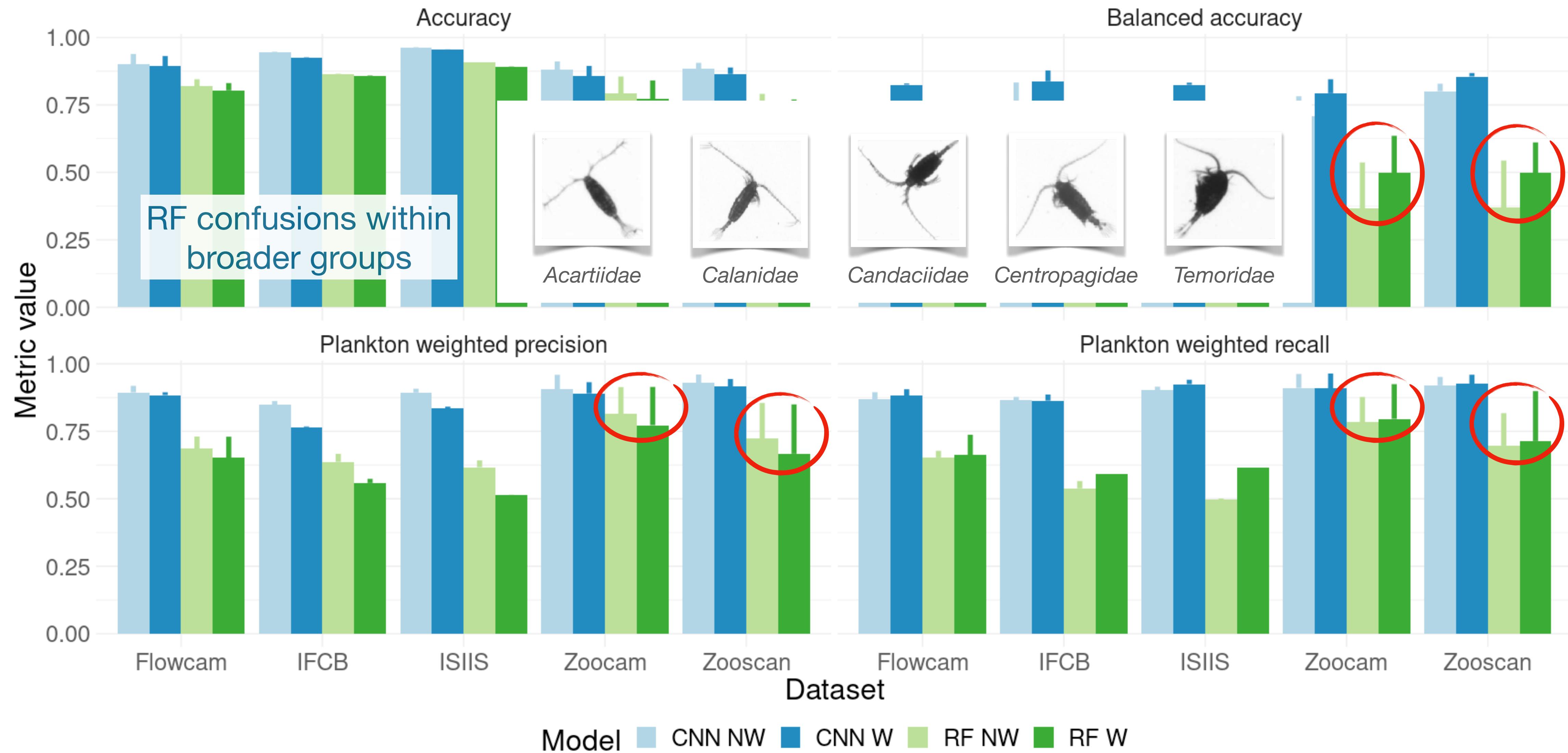
Small classes

- CNN >> RF
- W > NW



Results

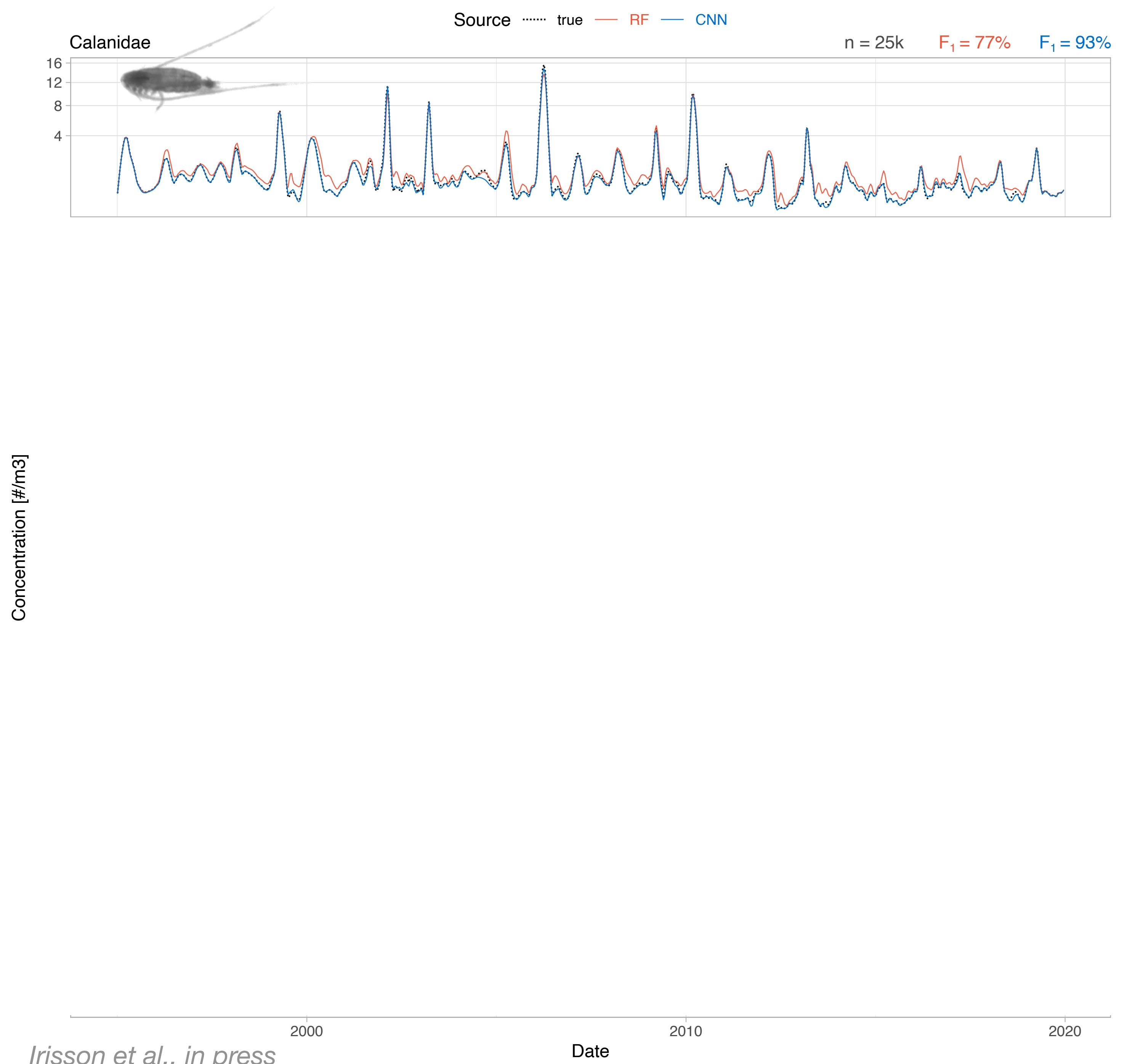
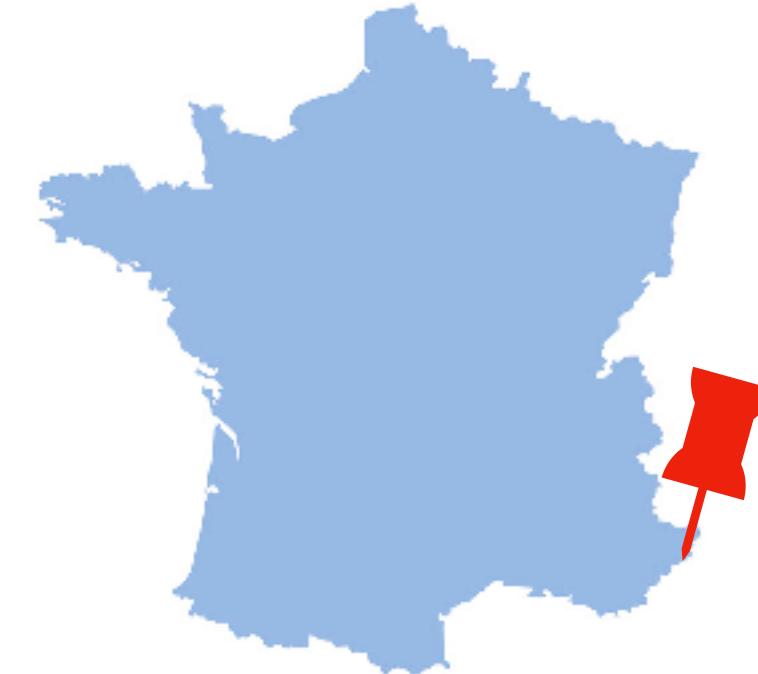
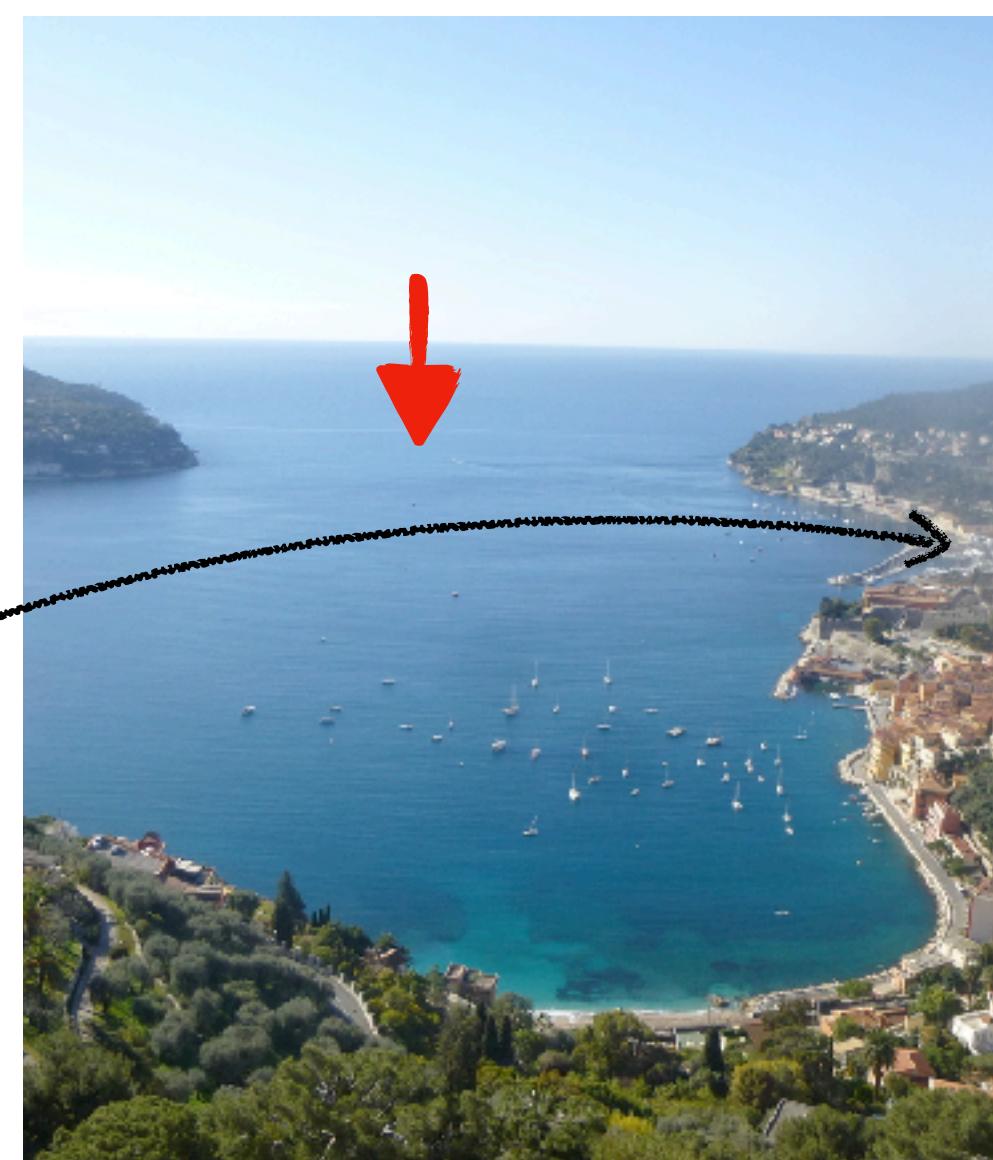
Model performances on detailed and broader classes



Application

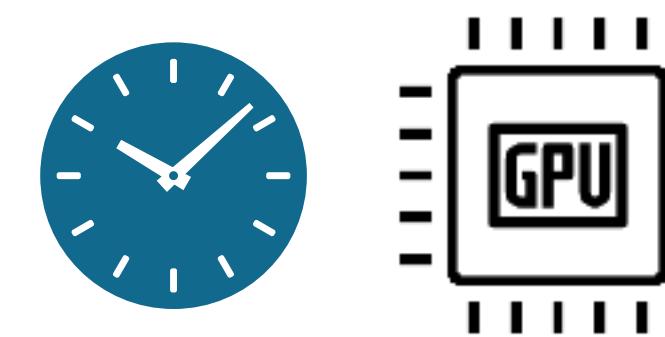


LOV LABORATOIRE
D'OCÉANOGRAPHIE
DE VILLEFRANCHE



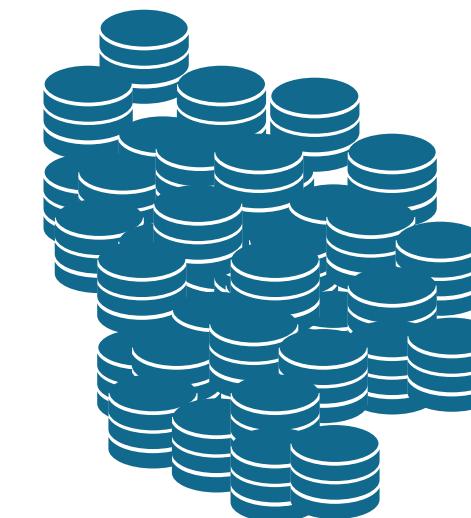
Discussion

- time and power consumption



- data availability

Lots lots lots lots lots lots of data

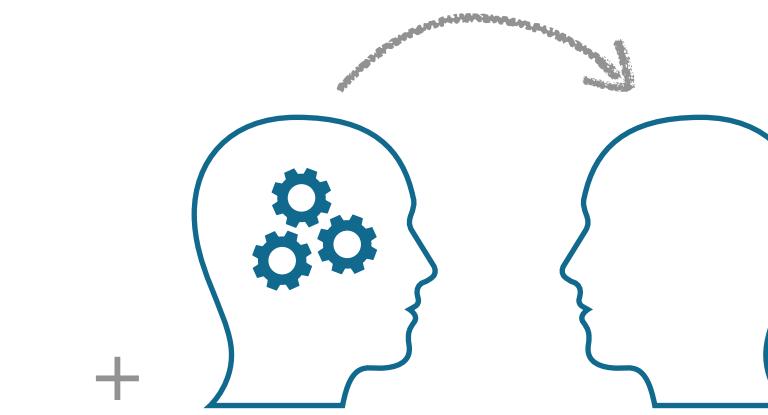


OR

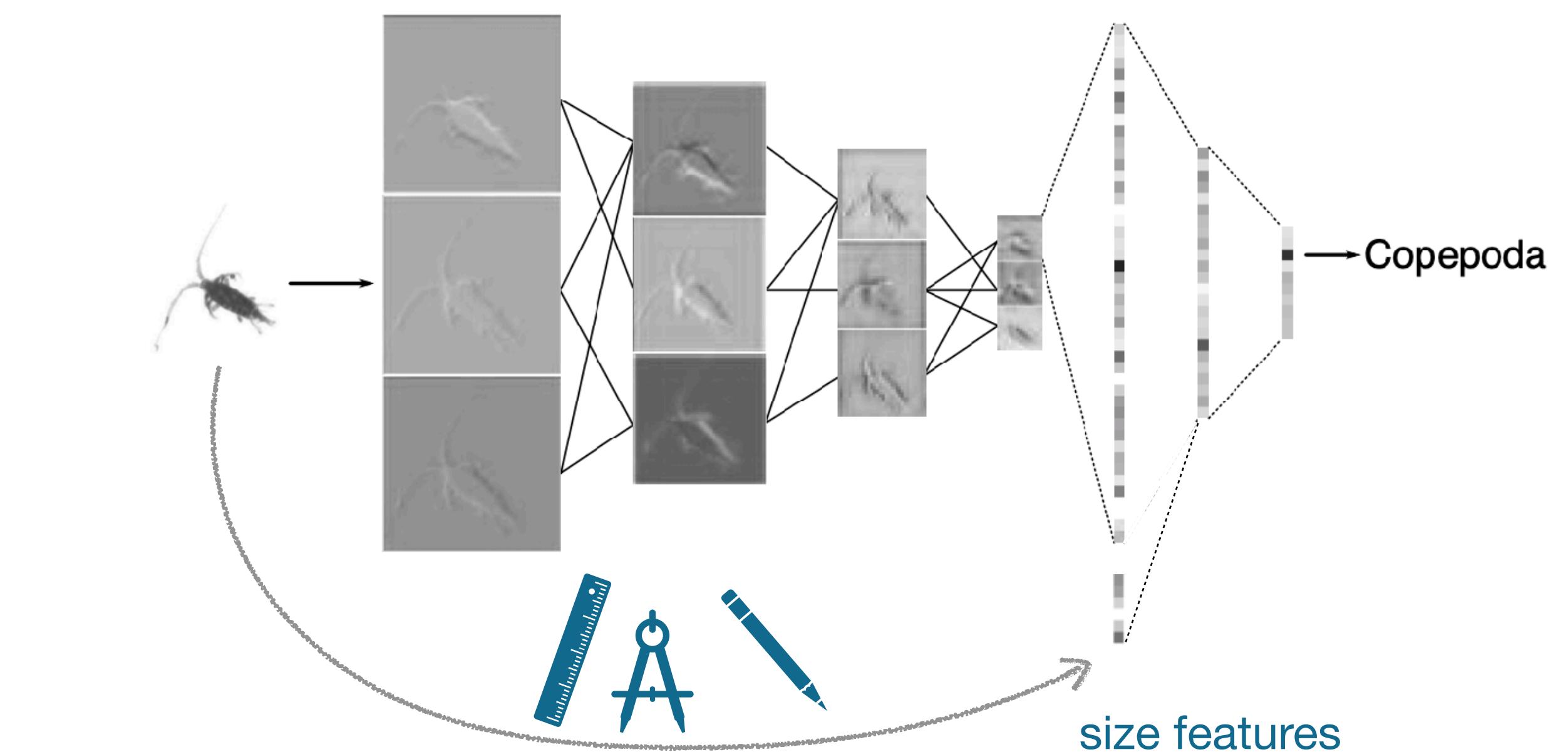
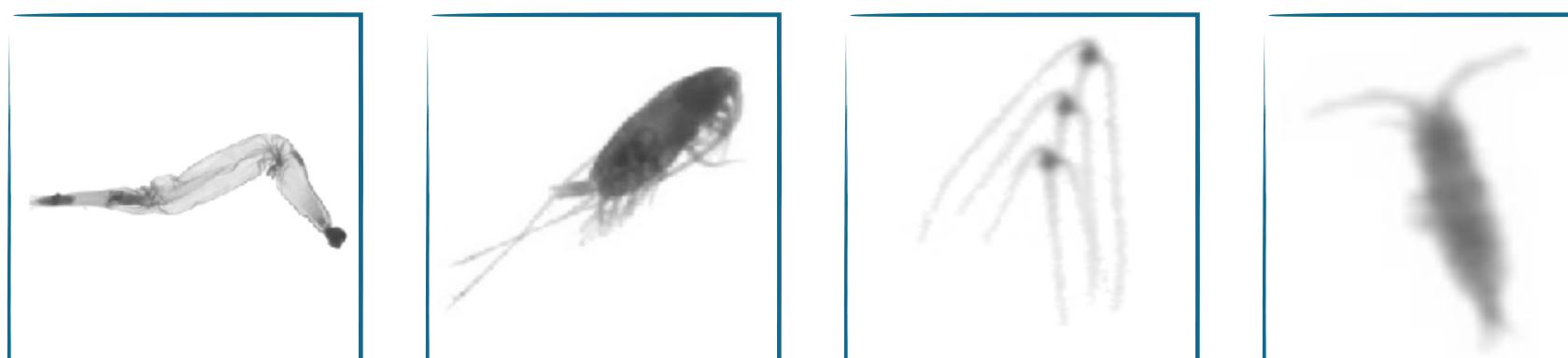
Lots of data



Transfer learning

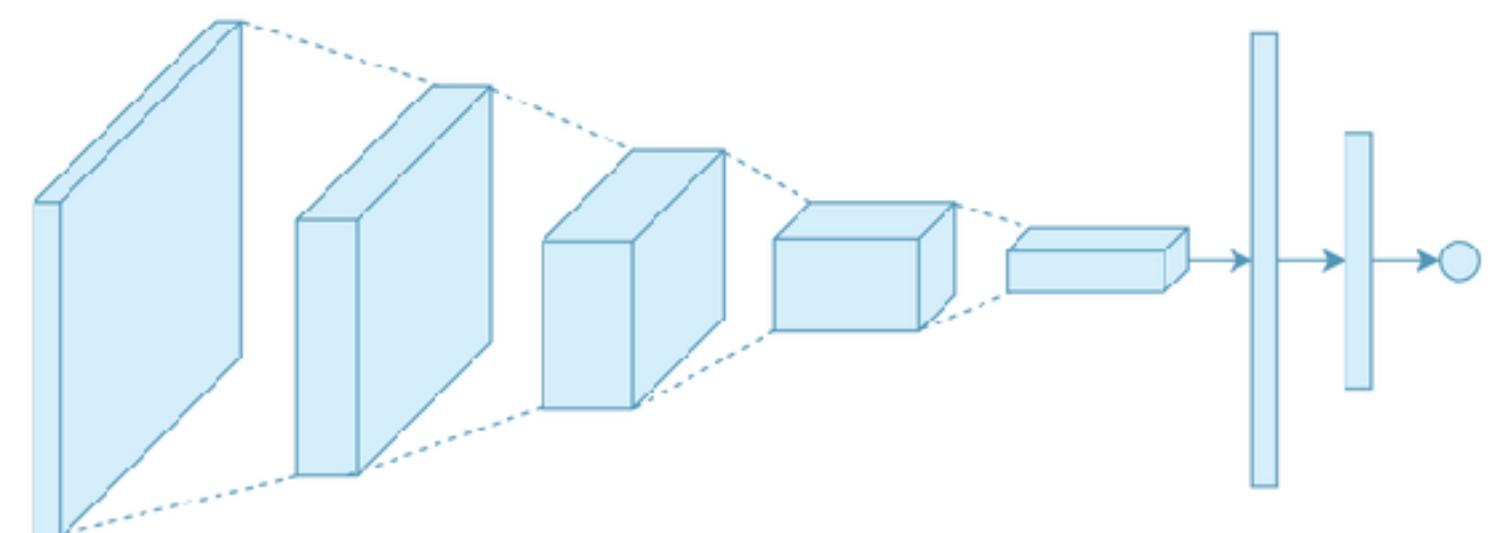


- objects size

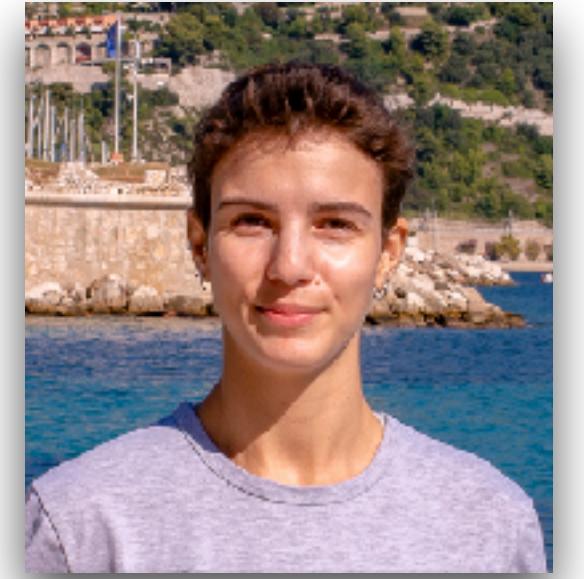


Conclusion

- Abundant plankton: RF ~ CNN
- CNN: more details on small classes
 - ▶ rare organisms
 - ▶ detailed taxonomy
 - ▶ low concentrations
- Manual validation



Thank you



Thelma Panaiotis
thelma.panaiotis@imev-mer.fr



https://github.com/ThelmaPana/plankton_classif_benchmark

