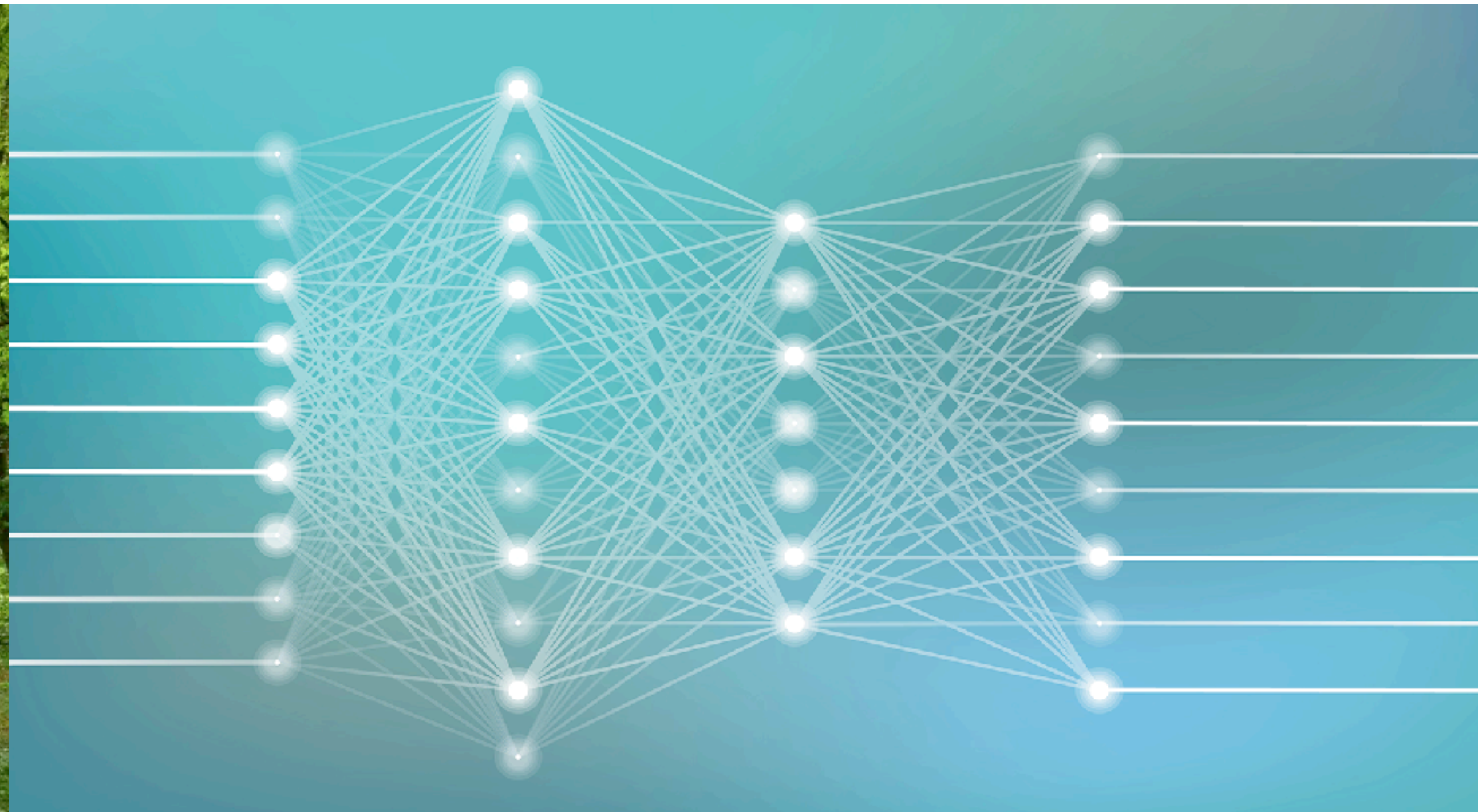
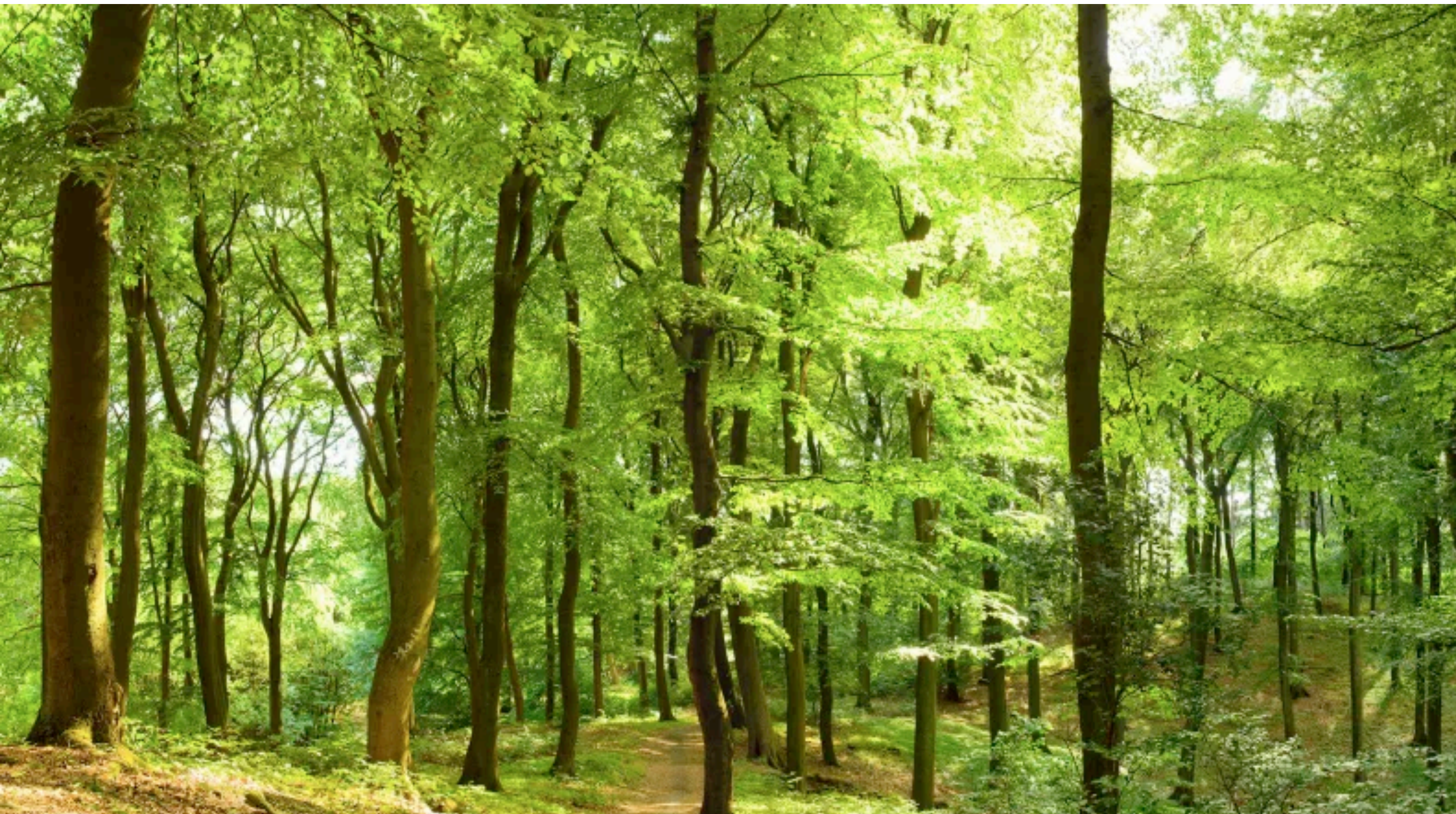


Benchmark of plankton image classification: CNN improve detection of rare taxa

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

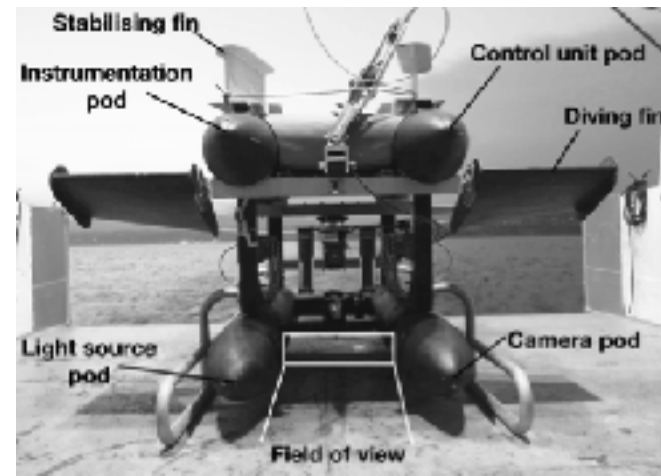
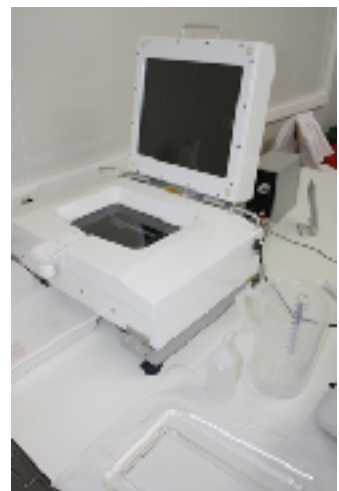
ASLO 2021 Aquatic Sciences Meeting



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

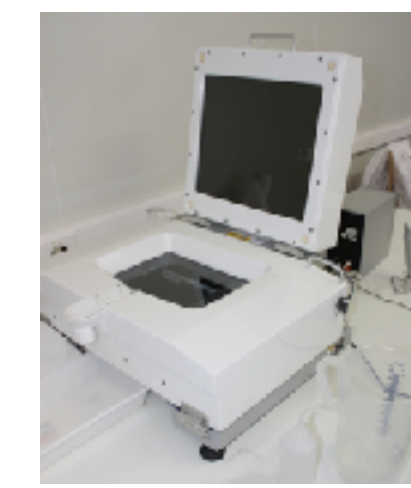


Lack of inter-comparison

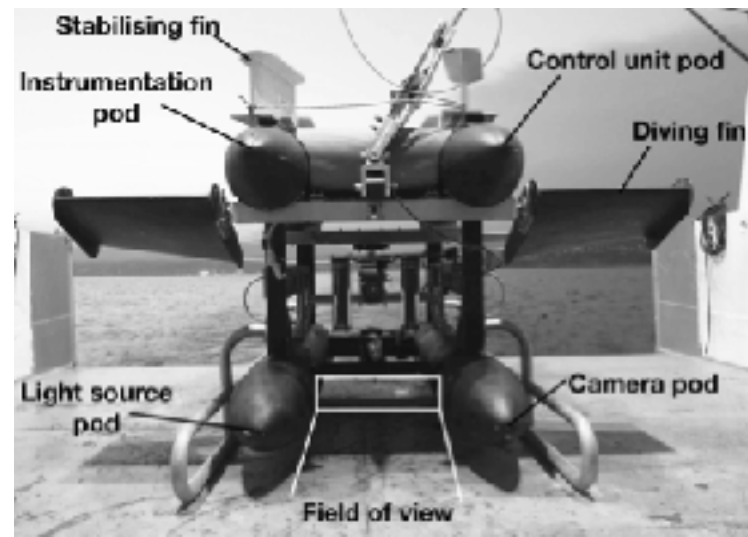


Material & methods

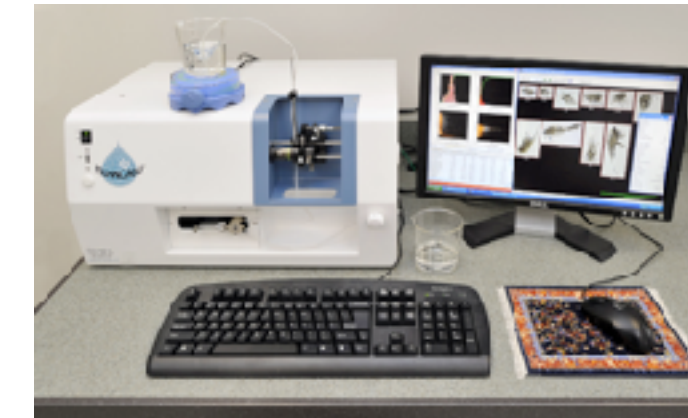
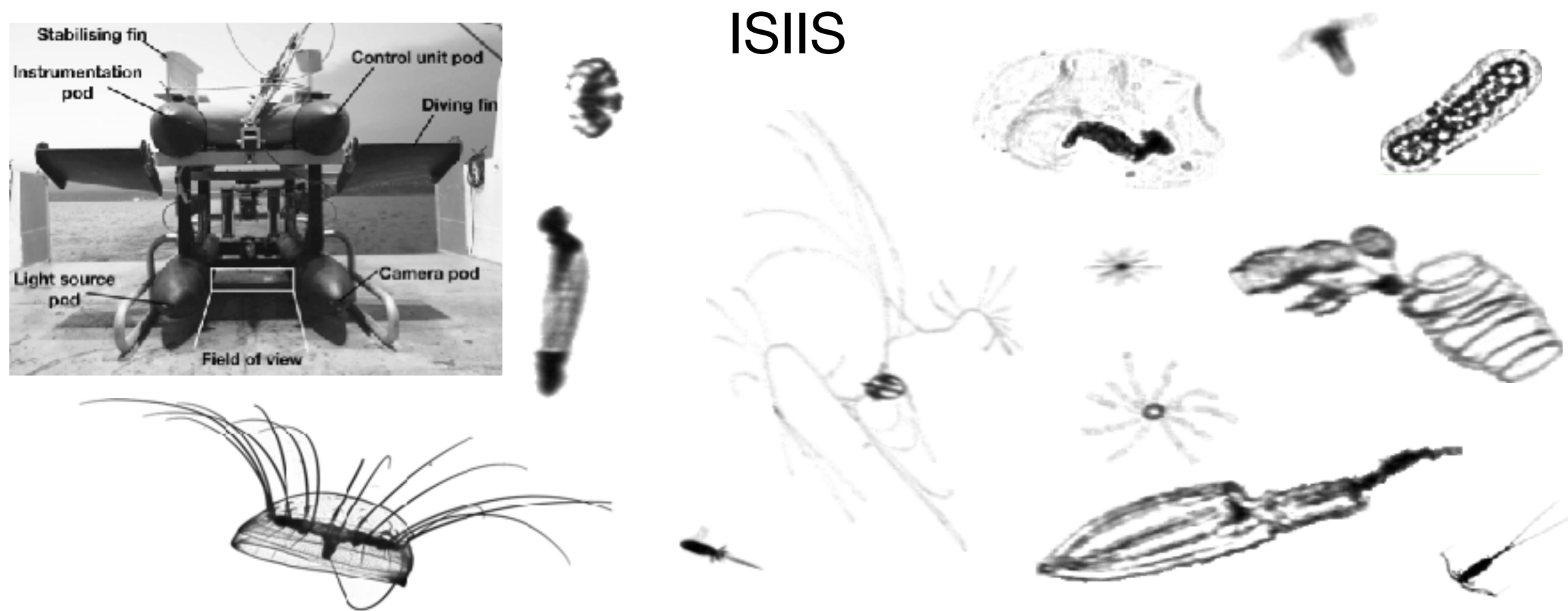
Datasets – imaging instruments



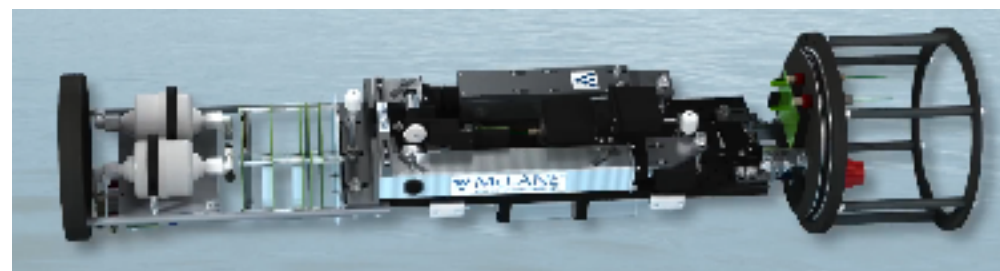
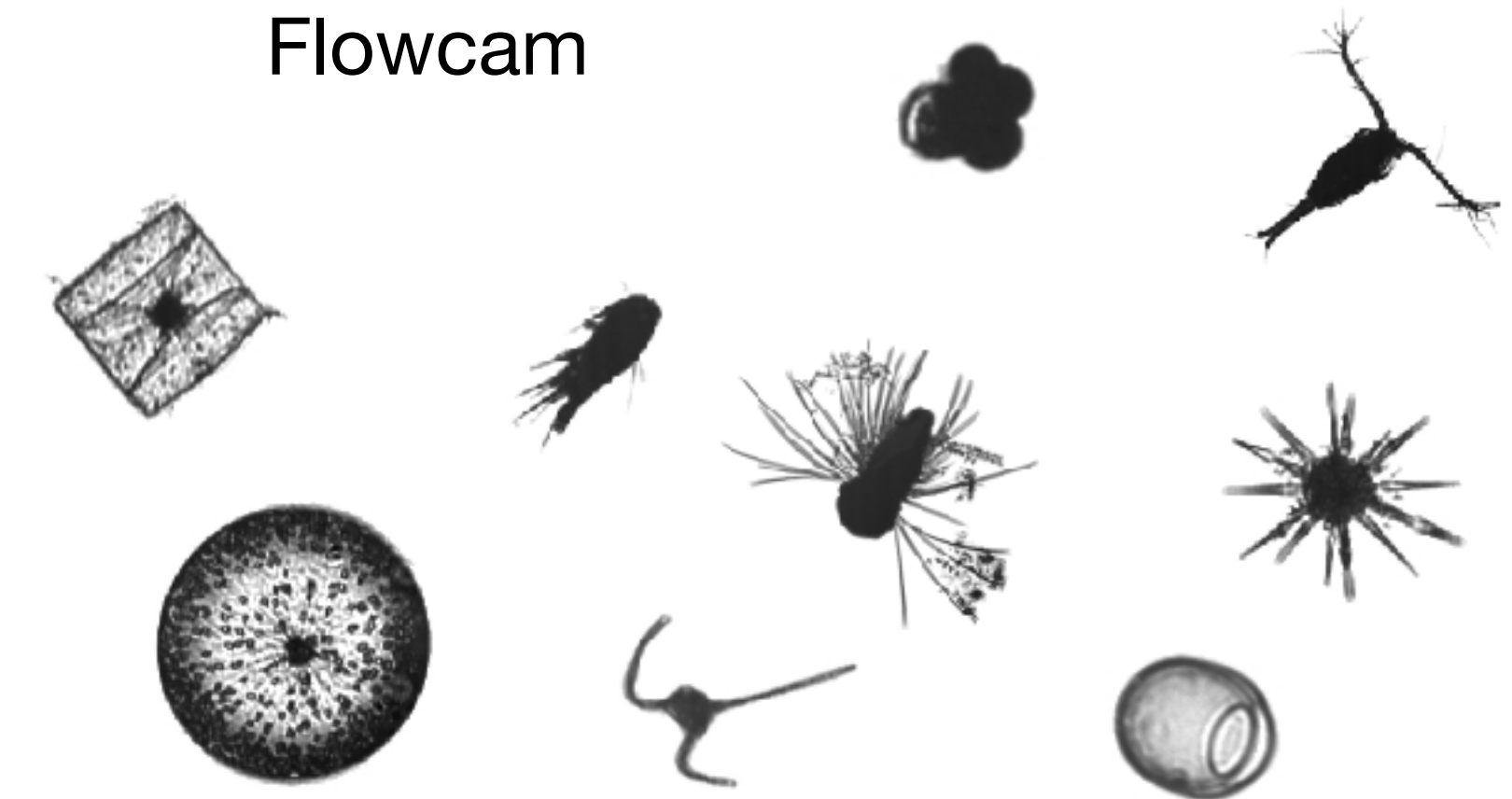
Zooscan



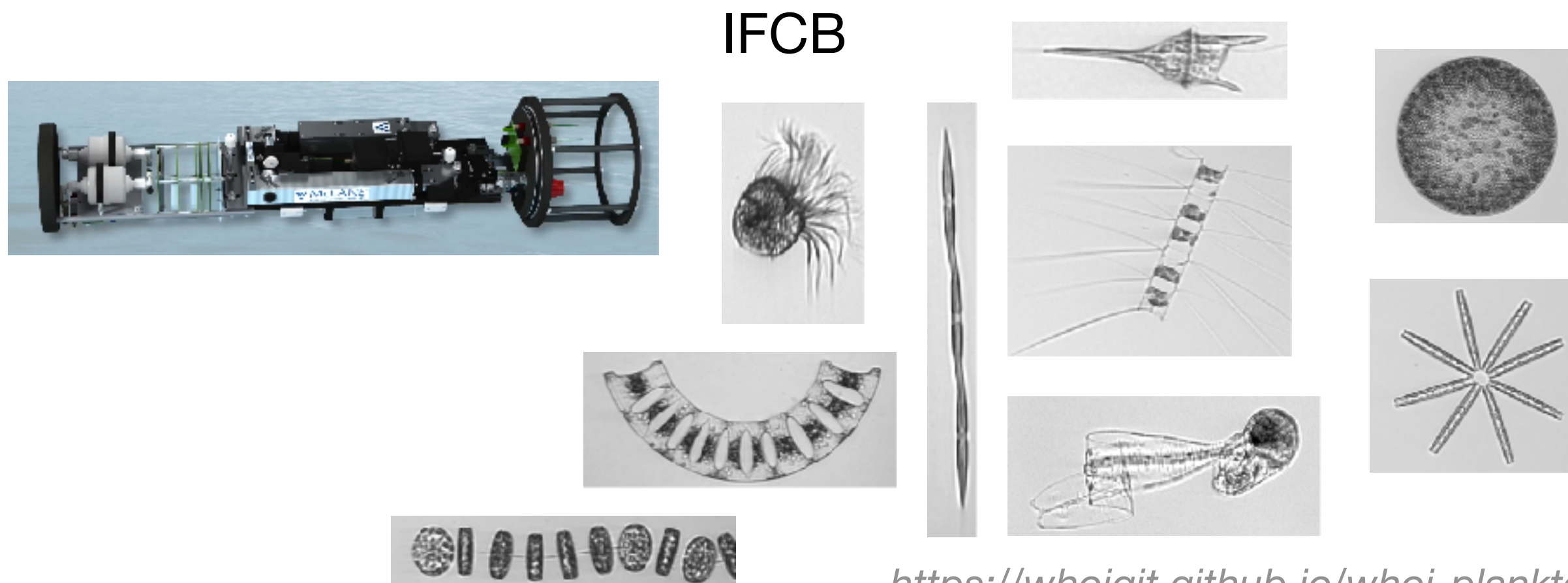
ISIIS



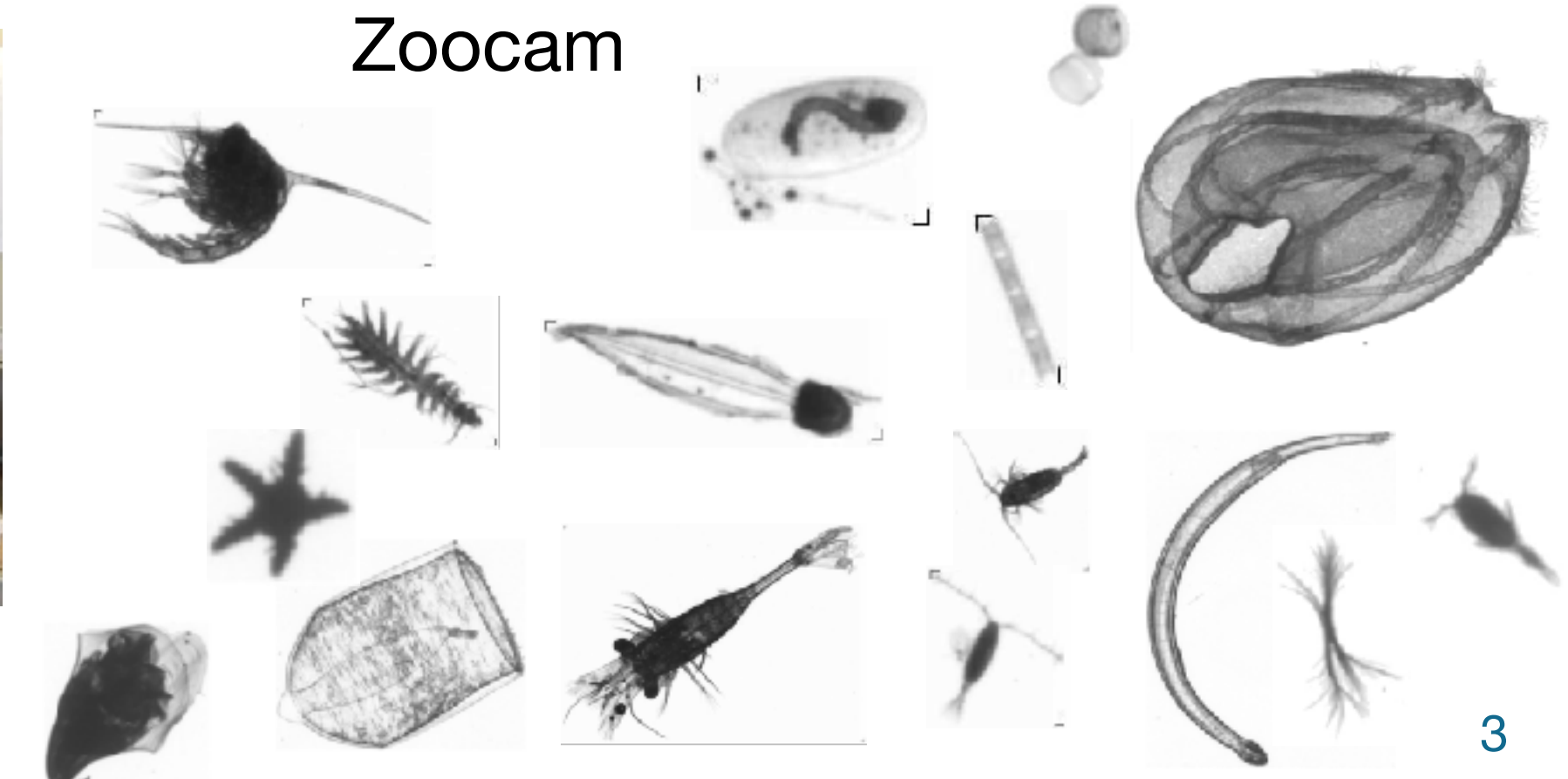
Flowcam



IFCB



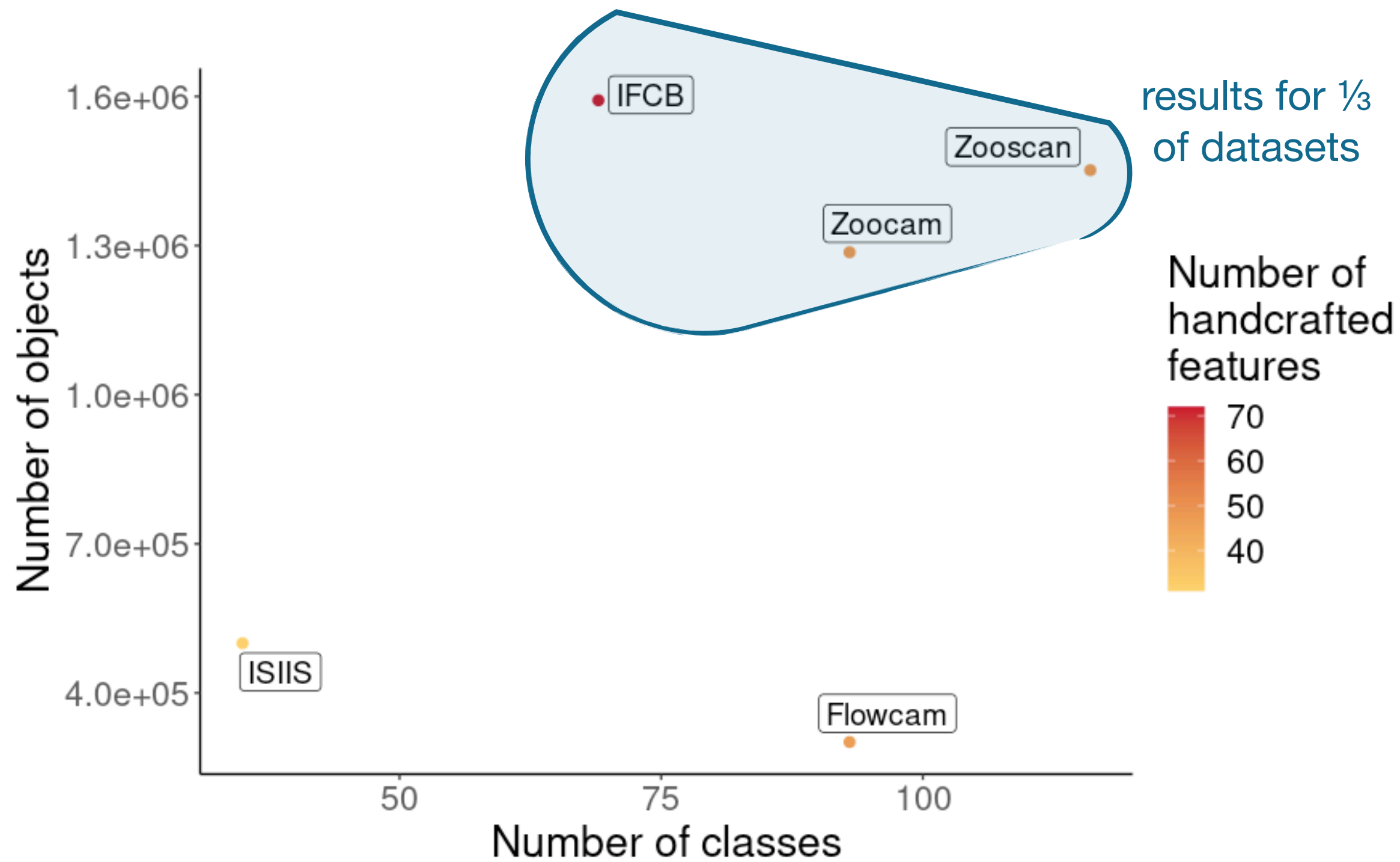
Zoocam



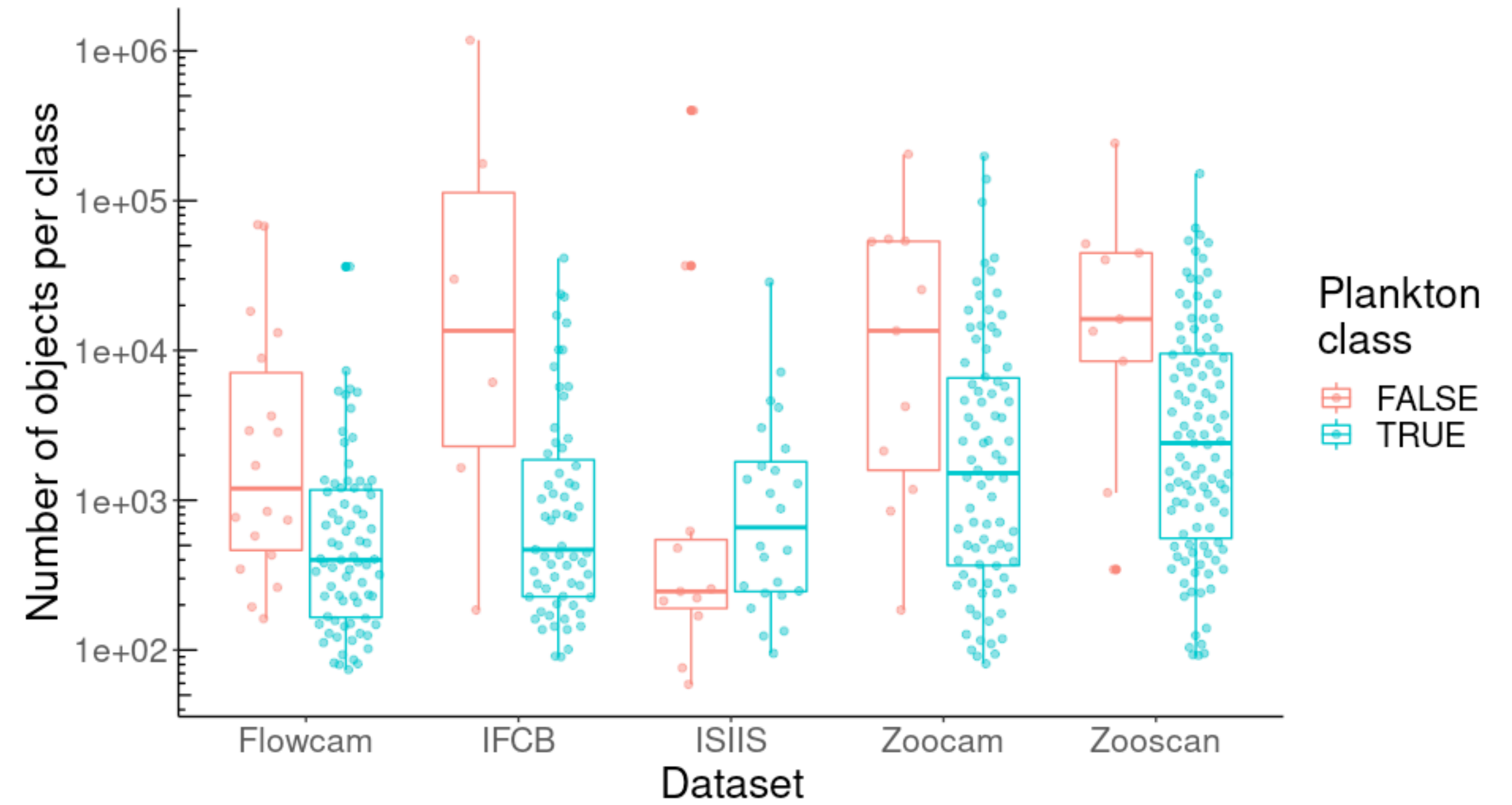
Material & methods

Datasets – composition

Number of objects and classes in each dataset



Distribution of objects per classes in each dataset



Strong unbalance



weight small classes

Class-stratified dataset split



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				
°--Oithonidae	Oithonidae	129334	Cyclopoida	139175
--oithon dae-bad'ocus	oithonidae-badfocus	4111	Cyclopoida	139175
--oithon dae-face-view	oithonidae-face-view	4569	Cyclopoida	139175
°--oithonidae-side-view-nc-ante	oithonidae-side-view-nc-antena	1161	Cyclopoida	139175
--dead	dead<Copepoda	17243	empty_Copepoda	17243
--Harpacticoida	Harpacticoida	645	Harpacticoida	645
--dead	dead<Harpacticoida	619	empty_Harpacticoida	619
--Euterpina	Euterpina	1449	Euterpina	1449
°--Microsetella	Microsetella	318	Microsetella	318
--multiple	multiple<Copepoda	5750	multiple_Copepoda	5750
°--Poecilostomatoida	Poecilostomatoida	679	Poecilostomatoida	692
--Corycaeidae	Corycaeidae	3659	Corycaeidae	3659
--Oncaeidae				
°--Oncaea	Oncaea	23336	Oncaea	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	667	part_Crustacea	1264
°--Hexapoda				
°--Insecta	Insecta	2	detritus	204132
--Bryozoa				
°--cyphonaute	cyphonaute	1593	cyphonaute	1593
--Chaetognatha	Chaetognatha	3096	Chaetognatha	3150
°--tail	tail<Chaetognatha	64	Chaetognatha	3150
--Chordata				
--Cephalochordata				
°--Branchiostomidae				
°--Branchiostoma	Branchiostoma	18	other_living	25454
--Cranialata				
°--Vertebrata				
°--Gnathostomata				
°--Actinopterygii	Actinopterygii	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 temp	41455	egg_Engraulidae	41455
--egg	egg<Actinopterygii	5138	egg_Actinopterygii	5138

Material & methods

Models – class weights

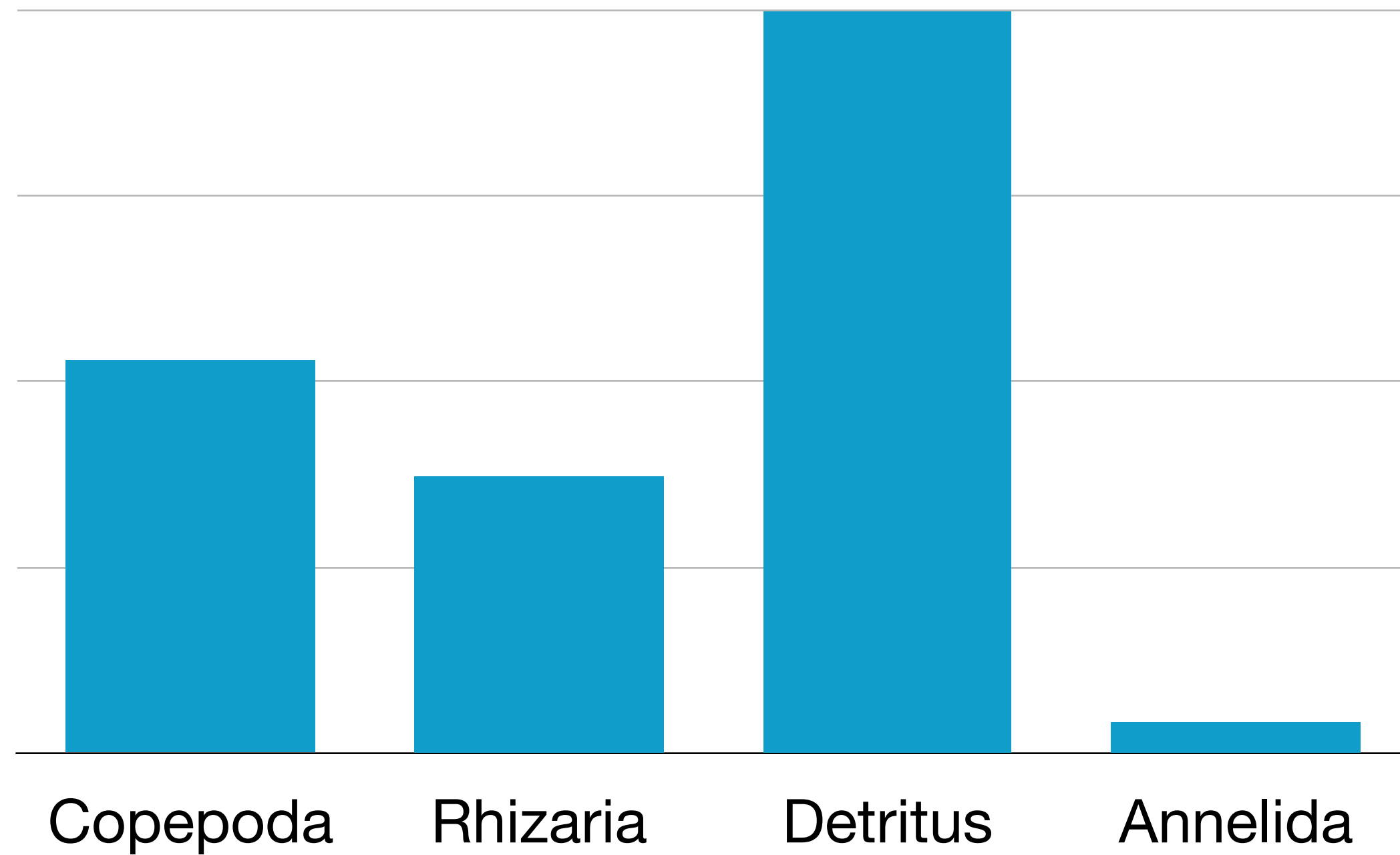
Weights for class i

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

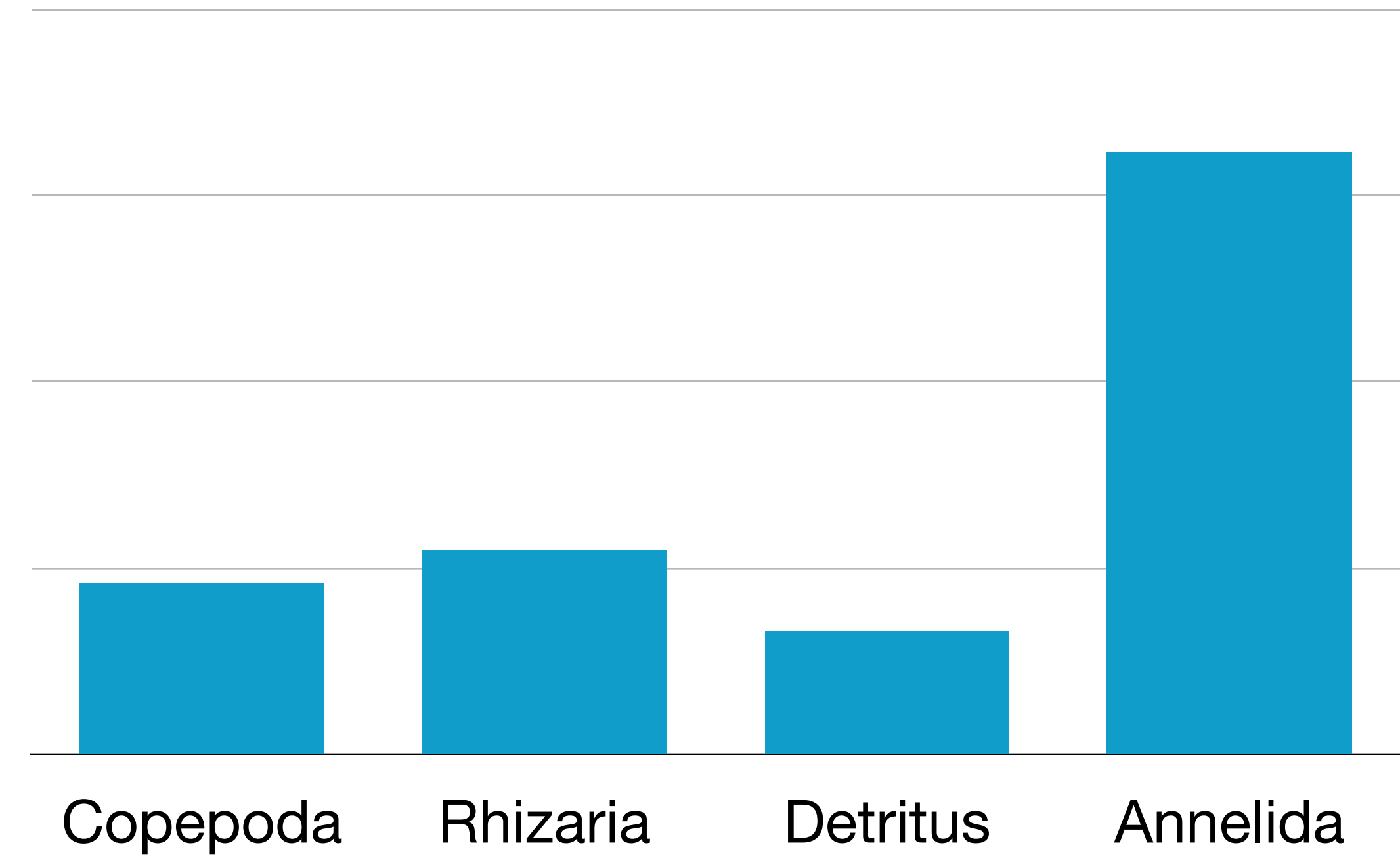
Number of objects in largest class

Number of objects in class i

Object number



Class weights

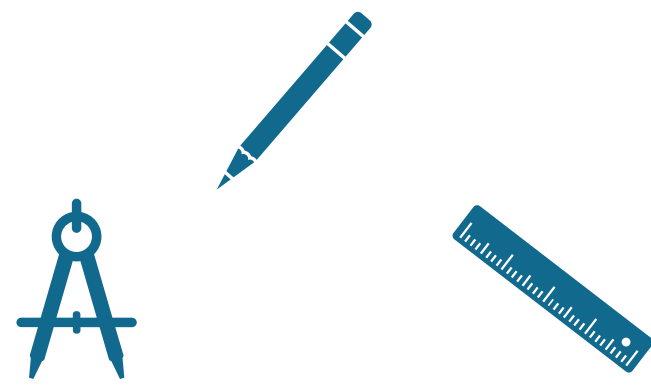


Material & methods

Models

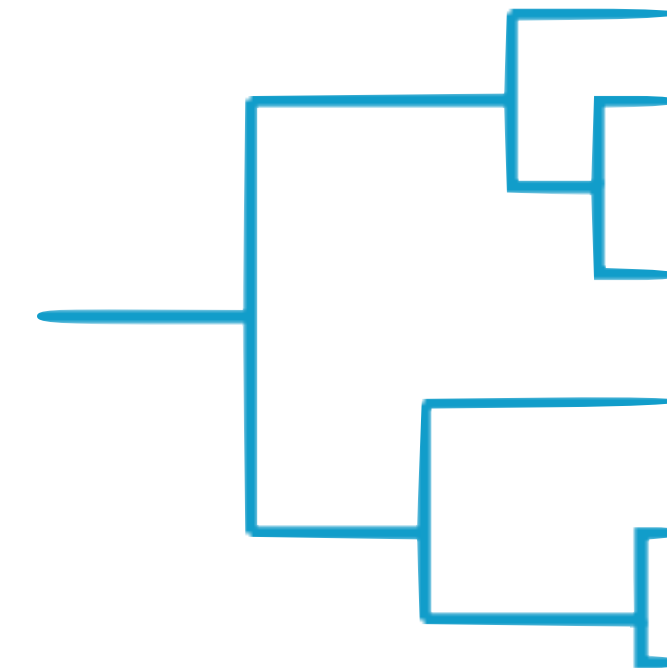
Image → Features → Classification

RF



- area
- gray levels
- feret
- circ
- ...

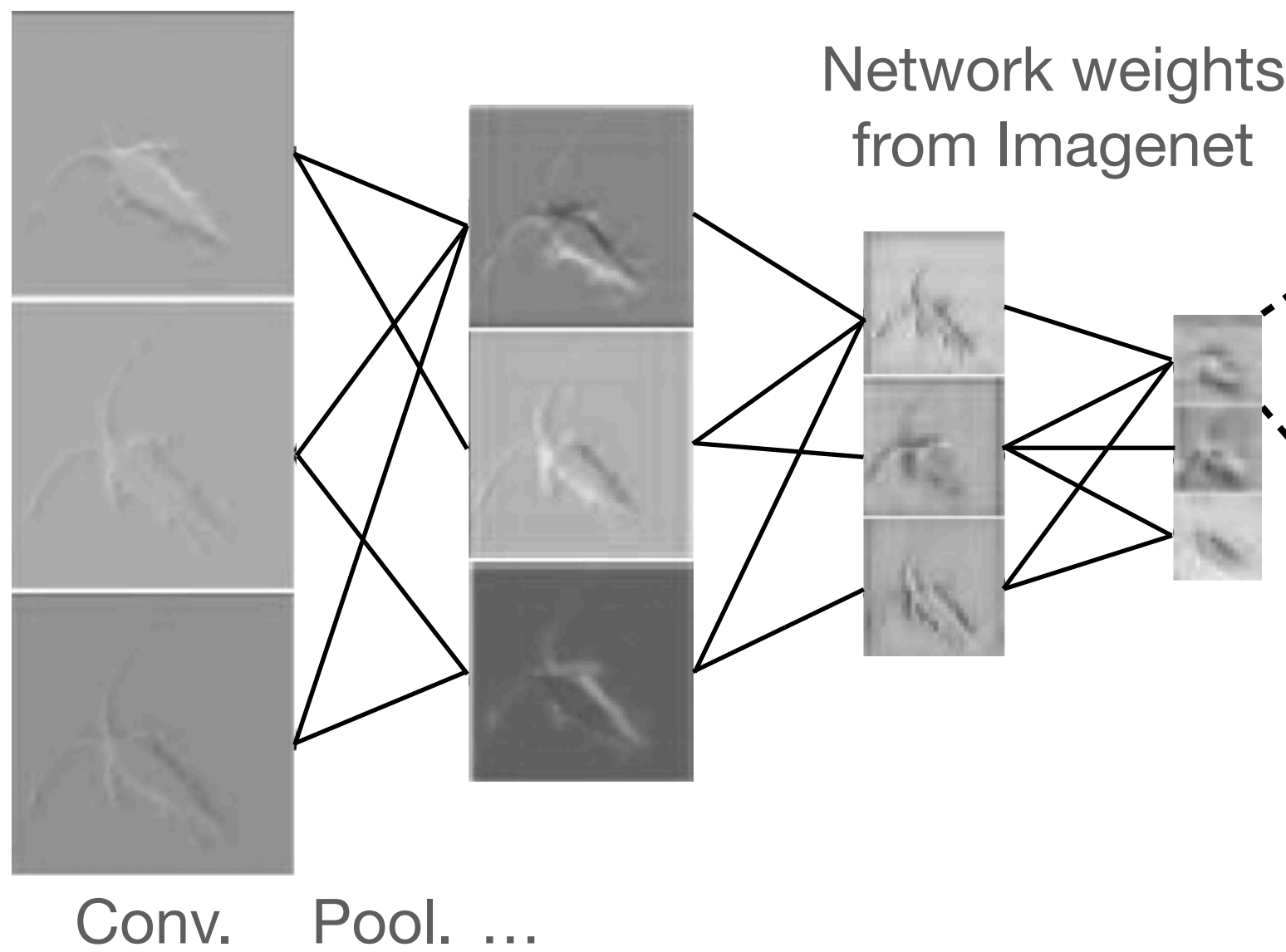
handcrafted features



Copepoda

Gridsearch for hyperparameters

CNN



Flatten

Fully connected layers & dropout

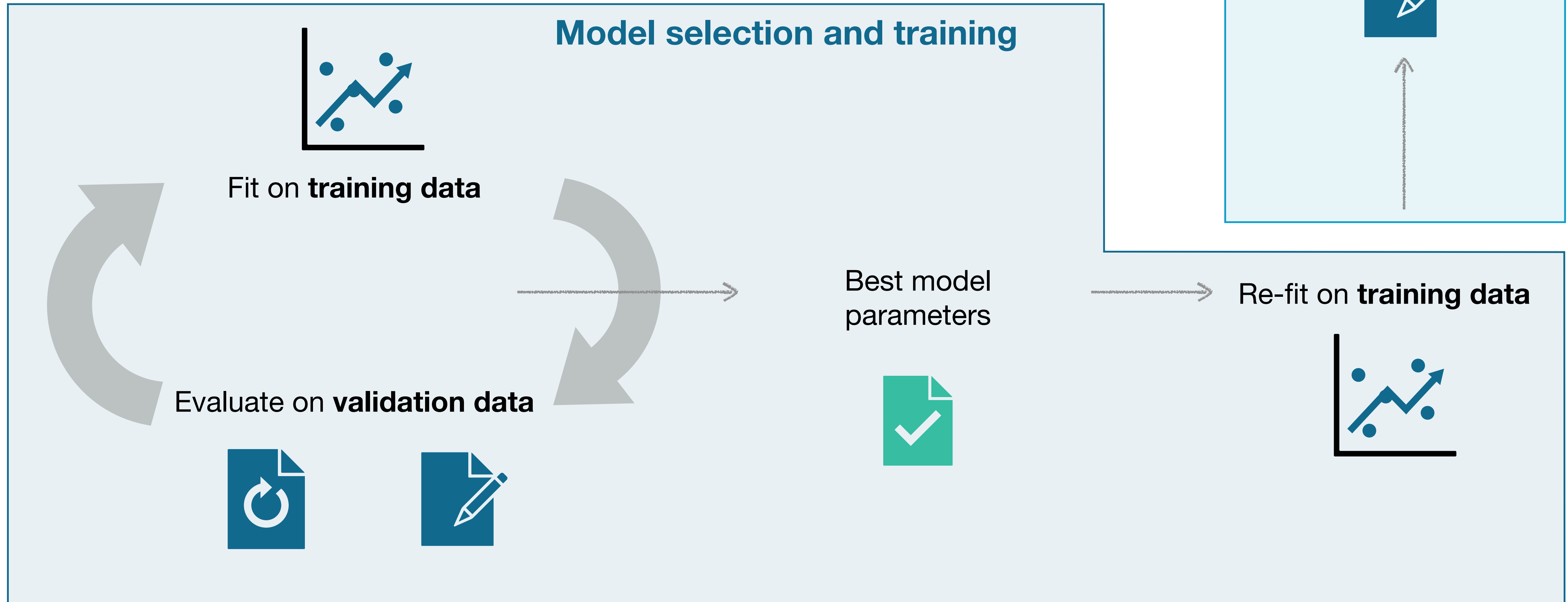
Classif.

CNN features

Copepoda

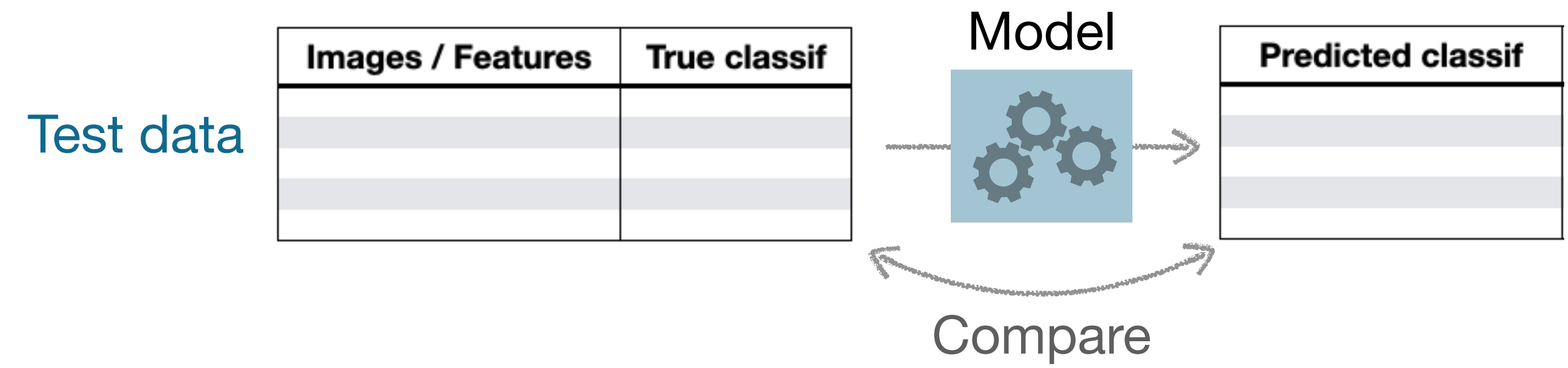
Material & methods

Models – training procedure



Material & methods

Model evaluation



Truth

	Egg	Copepoda	Detritus	
Egg				3/6
Copepoda				3/4
Detritus				
	3/4	3/5		

Prediction

Accuracy

$$\frac{\square}{\square + \square}$$

Balanced accuracy

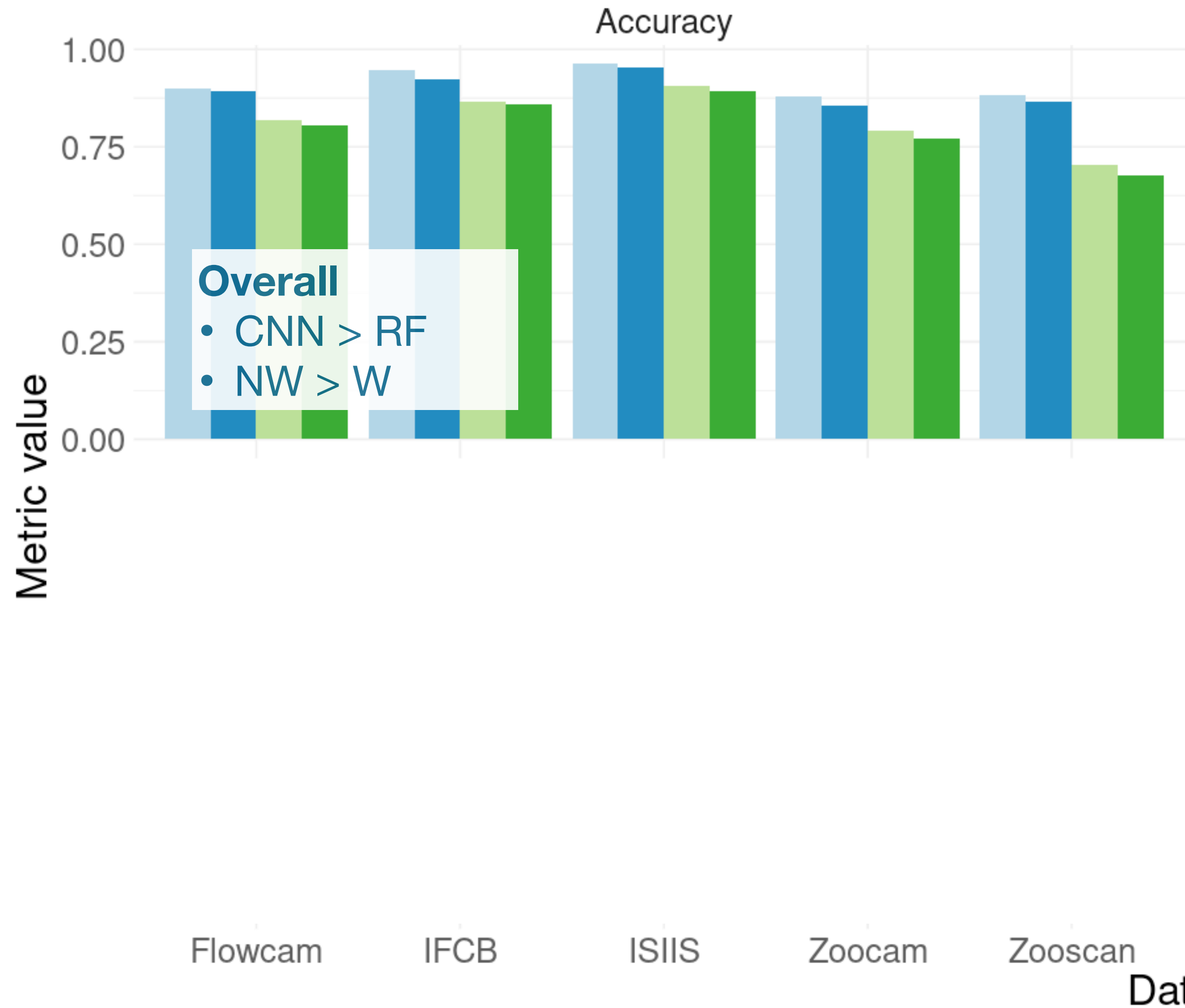
Weighted plankton precision 60%

Weighted plankton recall 67%

Results

CNN = convolutional neural network
 RF = random forest
 W = weighted
 NW = non weighted

Model performances on detailed classes



Small classes

- CNN >> RF
- W > NW

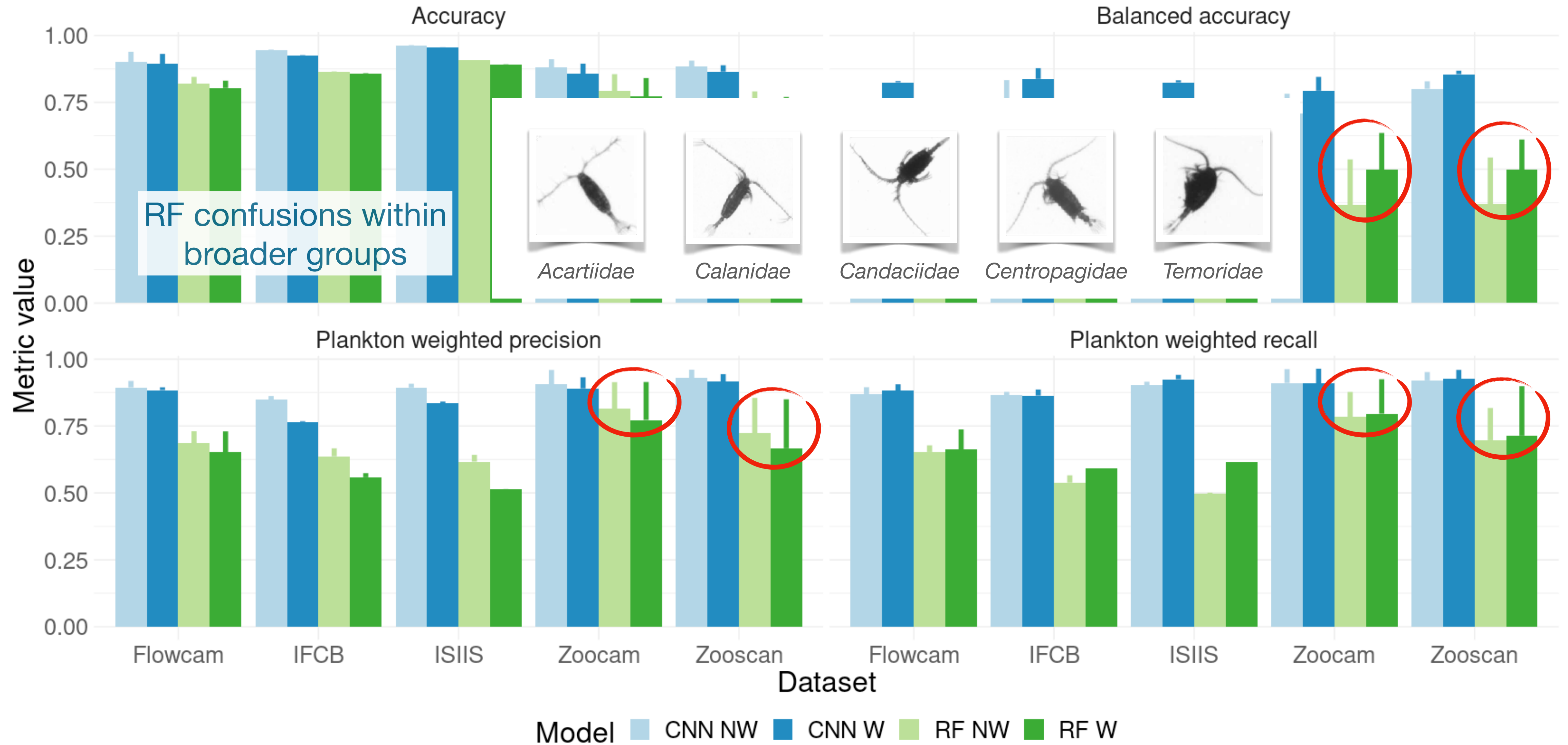
2%



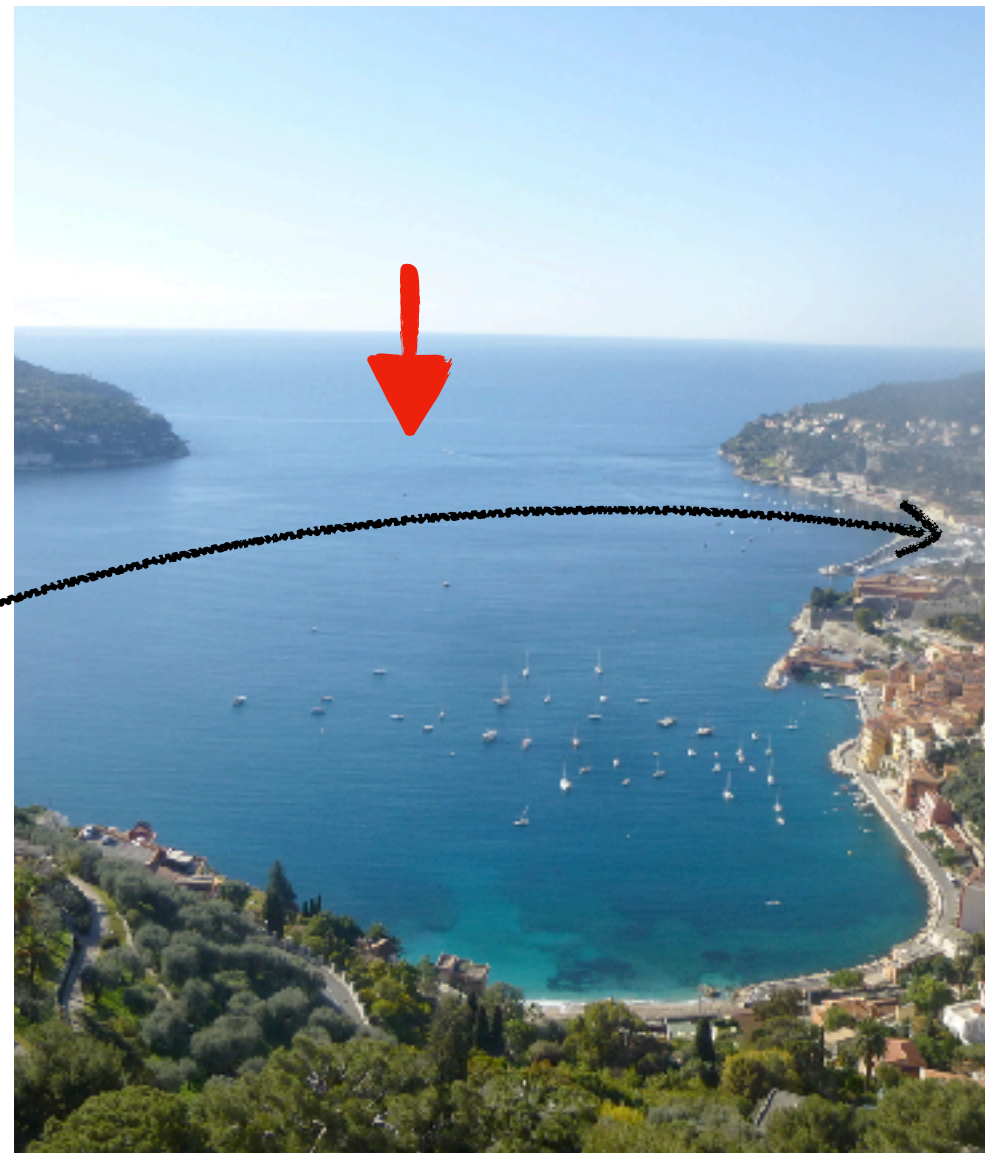
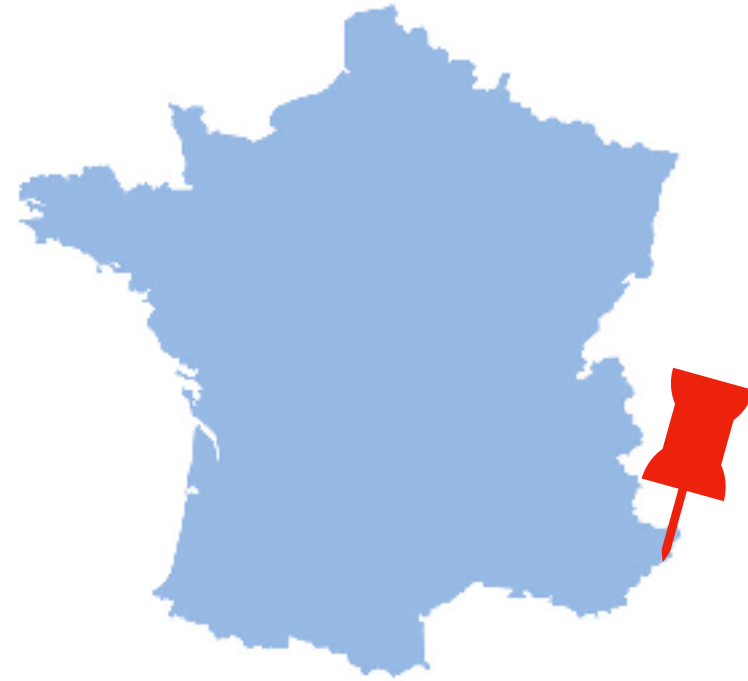
Model CNN NW CNN W RF NW RF W

Results

Model performances on detailed and broader classes

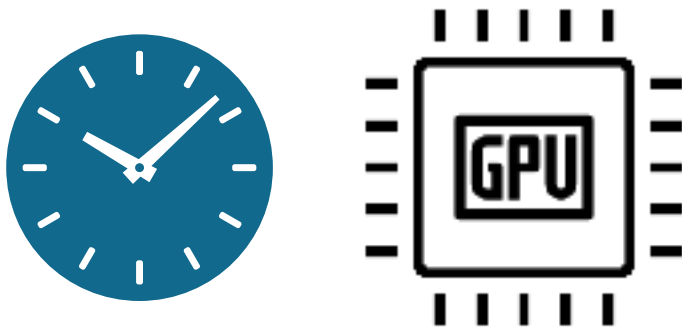


Application



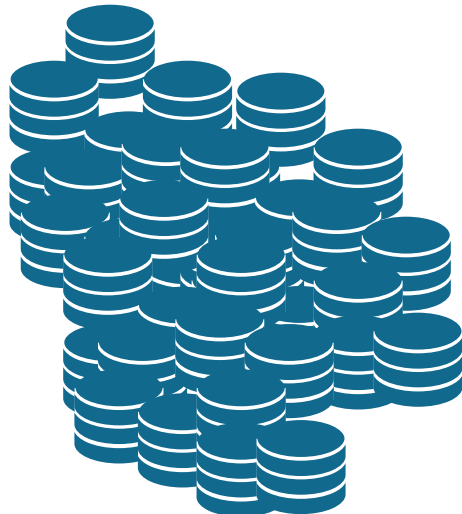
Discussion

- time and power consumption



- data availability

Lots lots lots lots lots lots of data

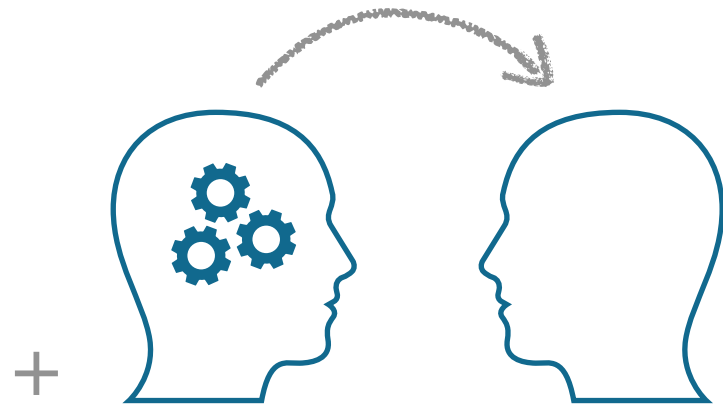


OR

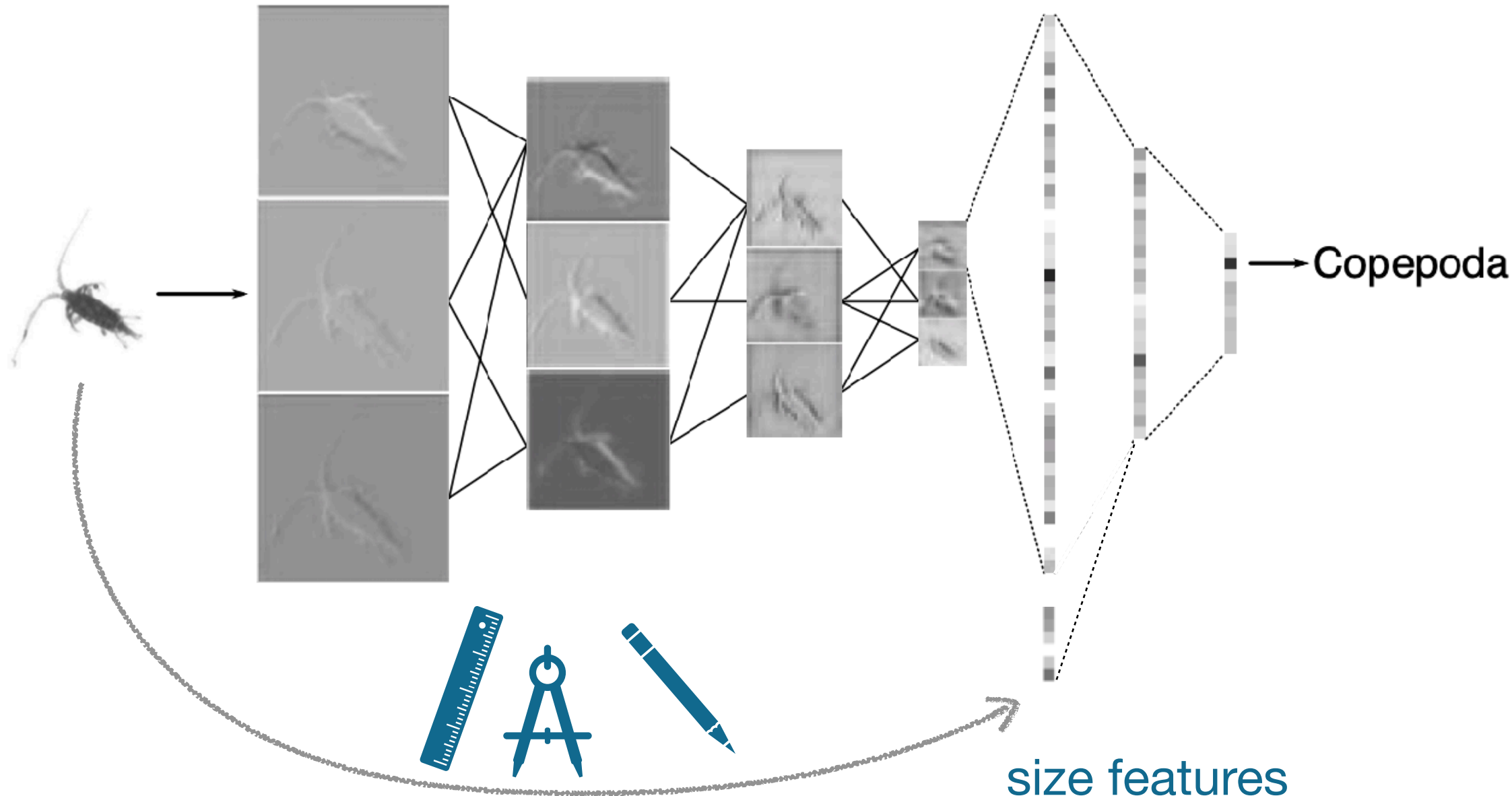
Lots of data



Transfer learning

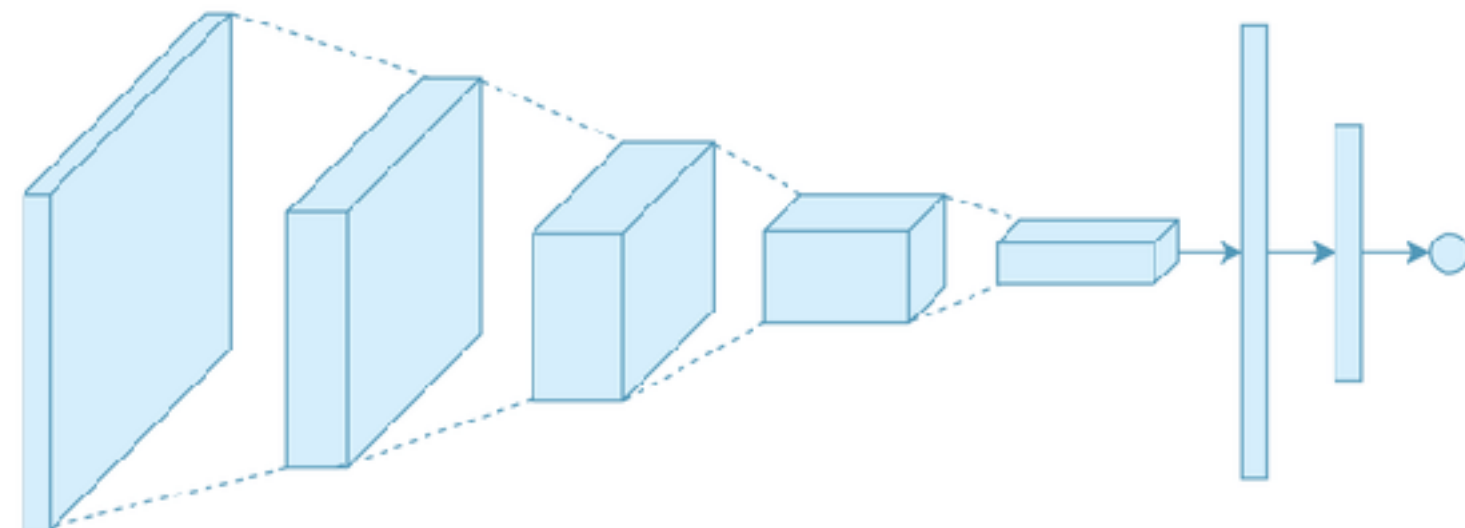


- objects size



Conclusion

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



Thank you



Thelma Panaïotis
thelma.panaïotis@imev-mer.fr



https://github.com/ThelmaPana/plankton_classif_benchmark

