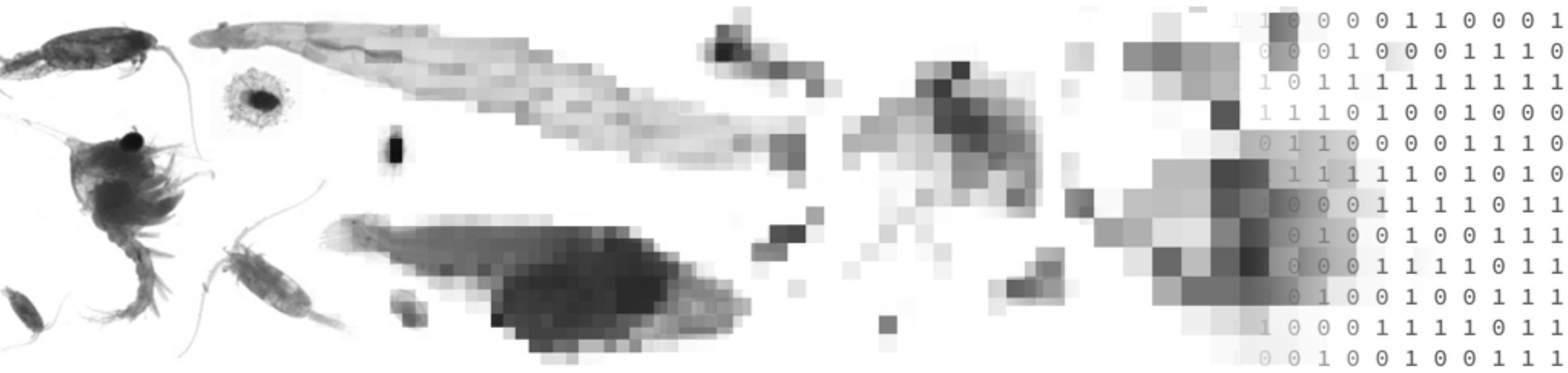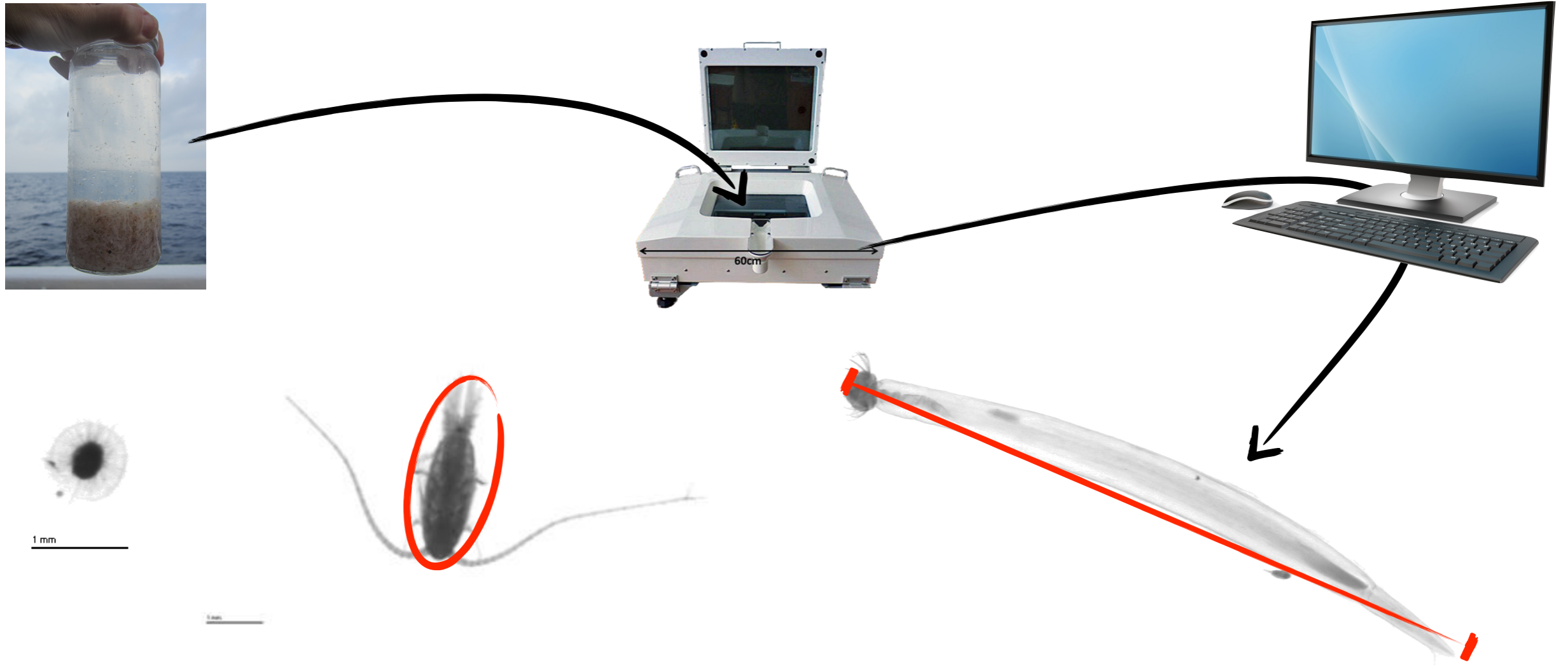SCIENCES
SORBONNE
UNIVERSITÉ

Quantitative image processing with machine learning

How to turn images into data

Automatic extraction of morphological features

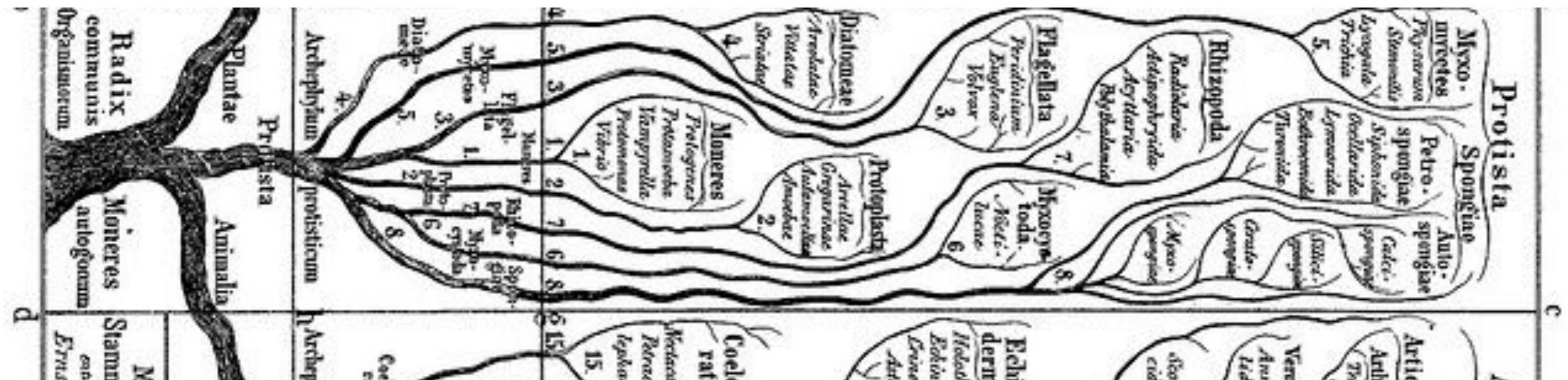|	feature 1	...	feature m
ind 1			
ind 2			
ind 3			
ind 4			

Current flow of images

ZooScan = 1 Bpx/y, UVP = 8.6Bpx/y, ISIIS=25Tpx/y
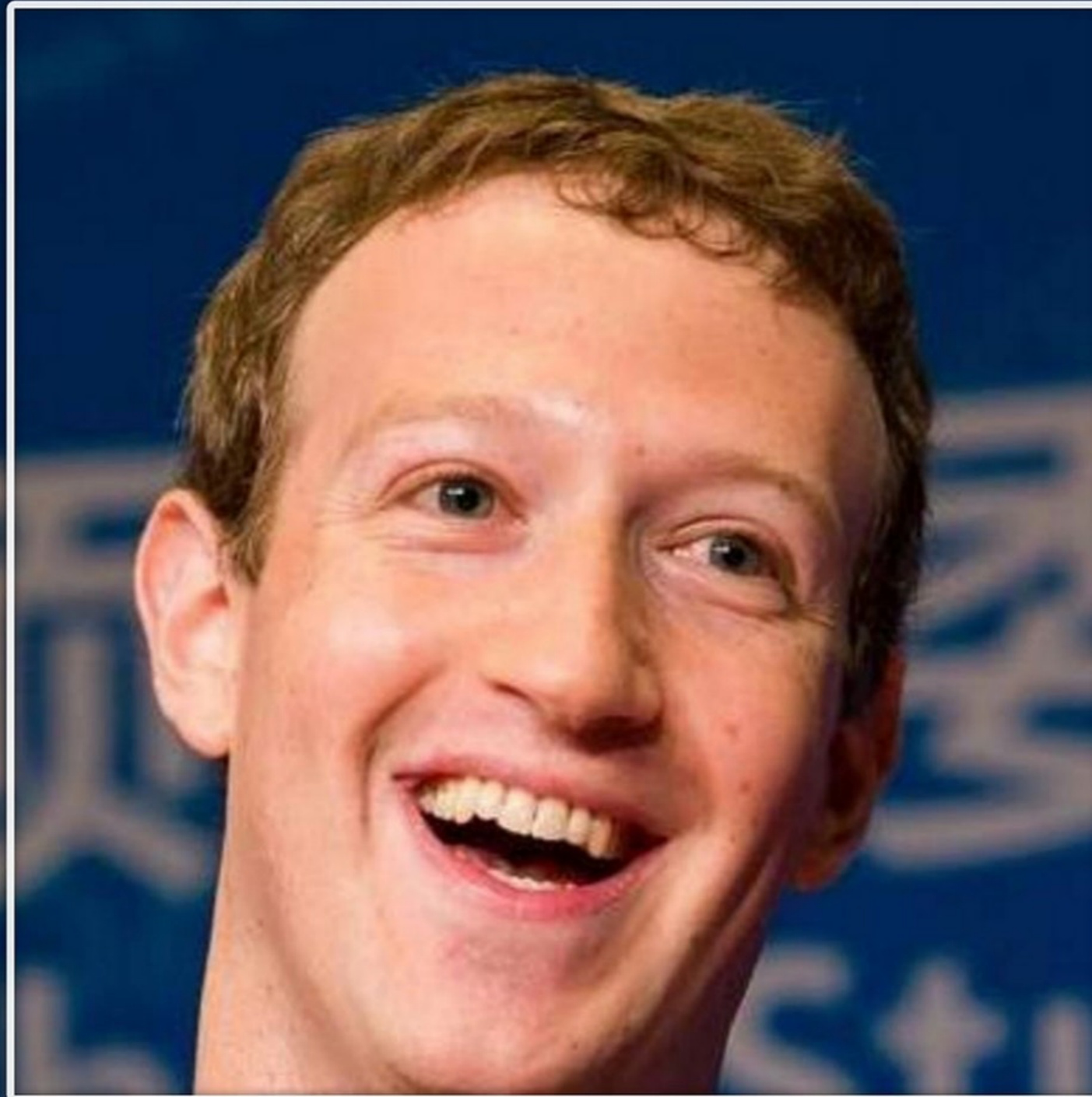⟹ Several million objects to classify per year

Classification of images

Because we like to put things in boxes

Demo

Deep Learning... how he does this:

A trendy topic

Works like your brain

input   retina   LGN   V1   V2   V3   LOC
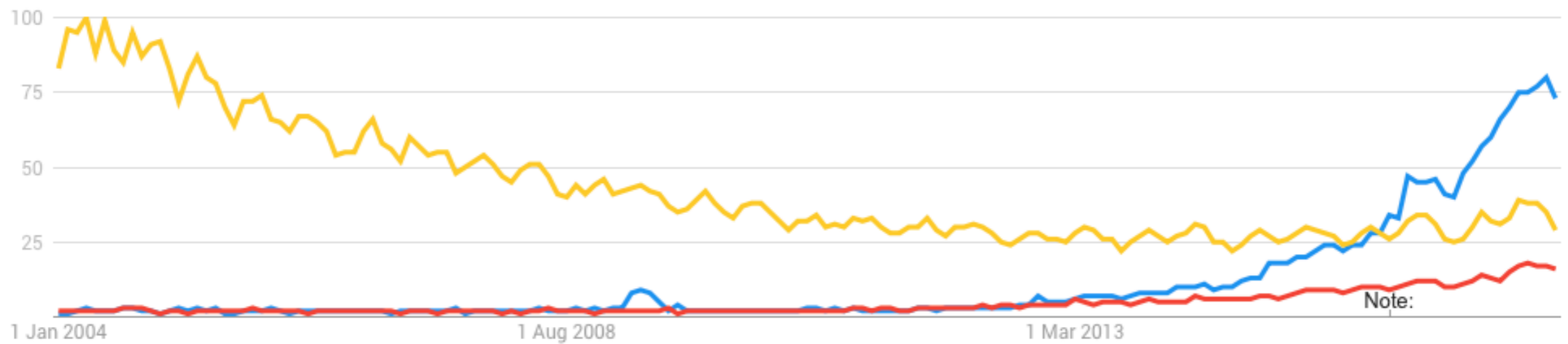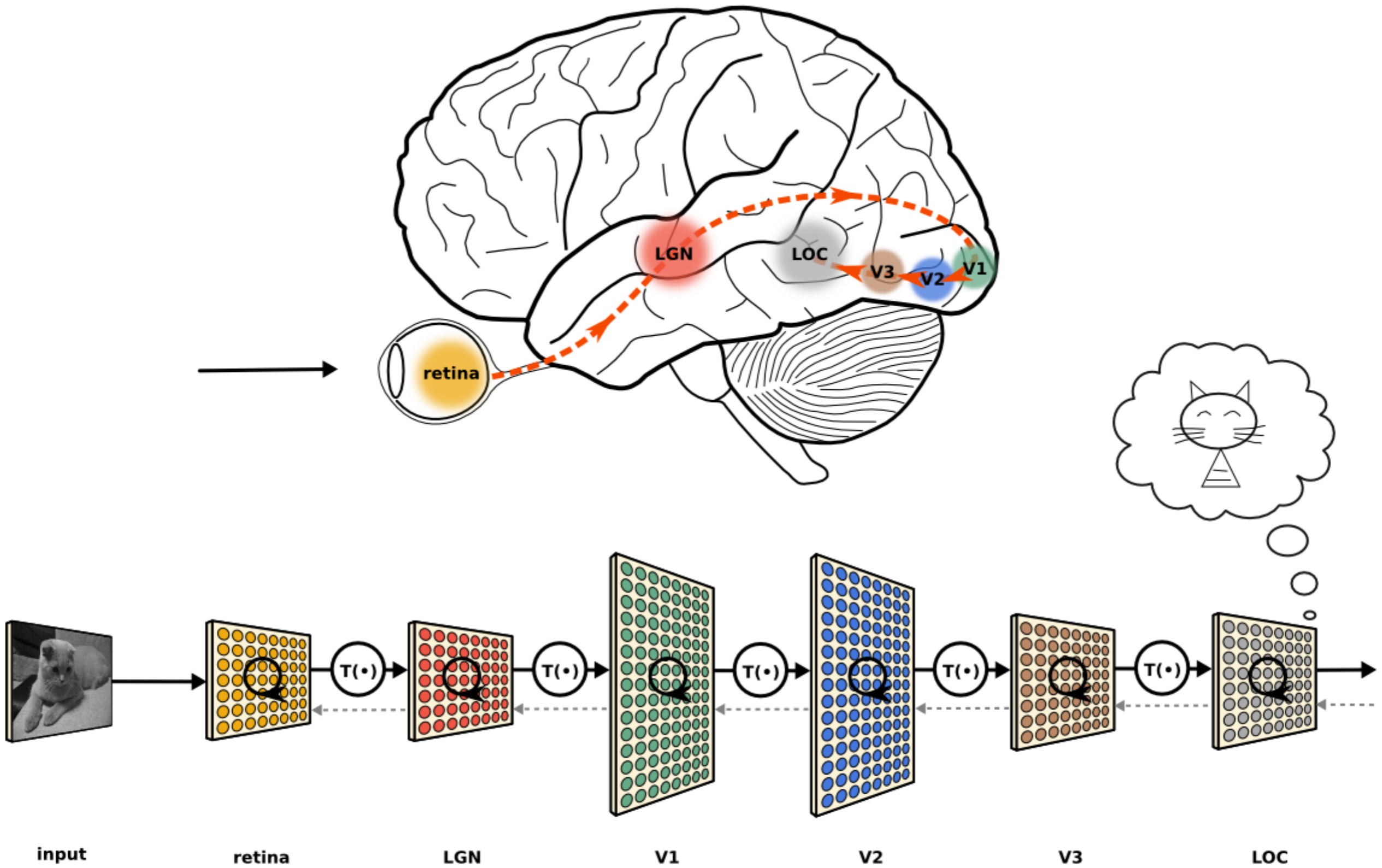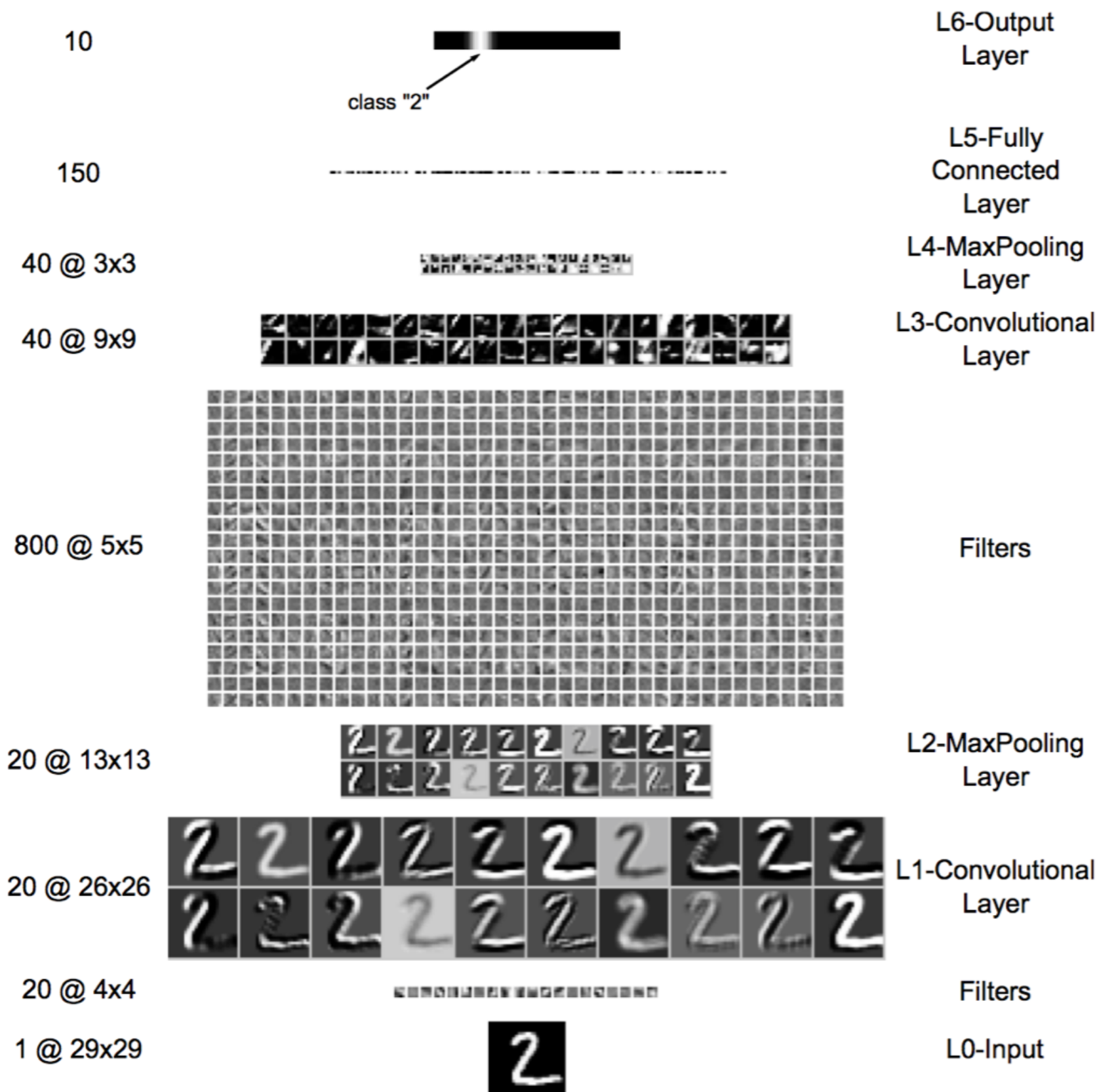
Inner workings of a CNN

CNN = Convolutional Neural Network

It extracts "morphological" **features**

(and classifies objects)

Features are "seen" at various scales

Training the network means assigning **weights**

10	L6-Output Layer
150	L5-Fully Connected Layer
40 @ 3x3	L4-MaxPooling Layer
40 @ 9x9	L3-Convolutional Layer
800 @ 5x5	Filters
20 @ 13x13	L2-MaxPooling Layer
20 @ 26x26	L1-Convolutional Layer
20 @ 4x4	Filters
1 @ 29x29	L0-Input

class "2"

+rectifiers = neuron activation functions
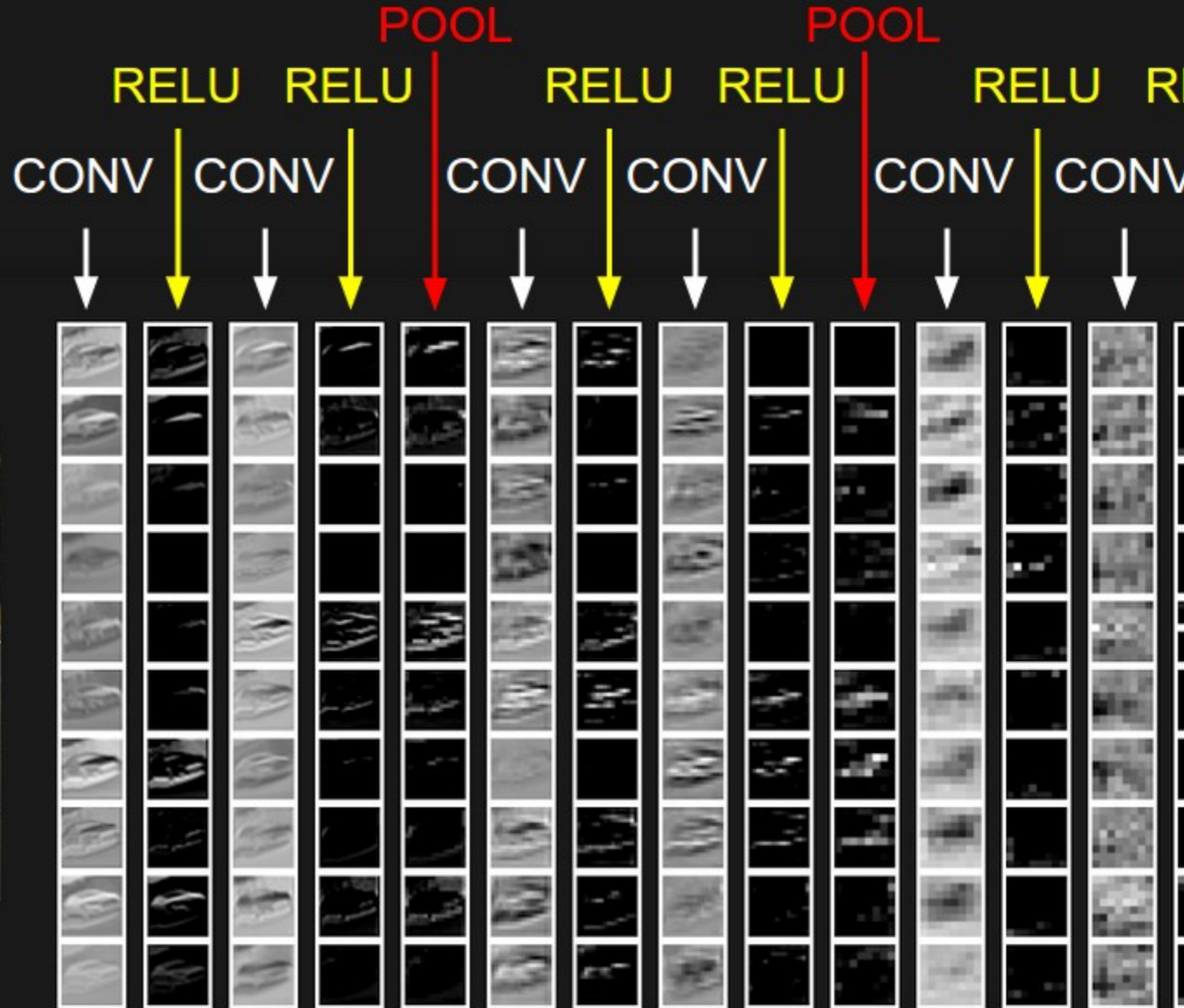
Deep Learning for plankton: Kaggle 2015

International competition for the classification of **plankton** images

60k images to classify in ~**120 groups** from a training set of 30k

1049 teams for a prize of $150k

Top 10 teams all used **CNNs**

83 to **85%** accuracy
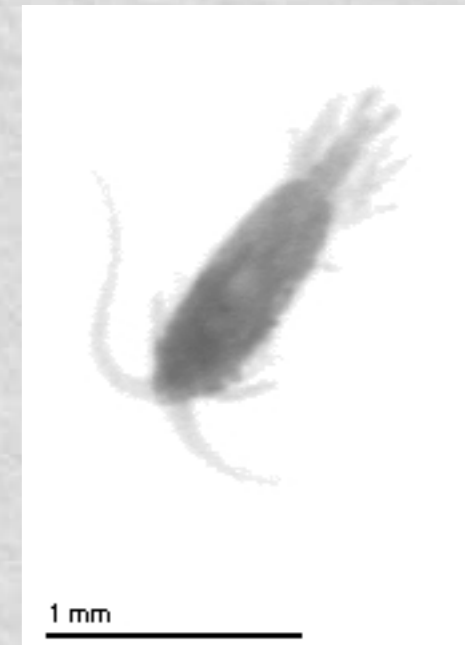
SparseConvNet in 3rd place

Why SparseConvNet?

Sparsity

   faster, more efficient

   varying input size

Factional Max Pooling

https://github.com/facebookresearch/
SparseConvNet
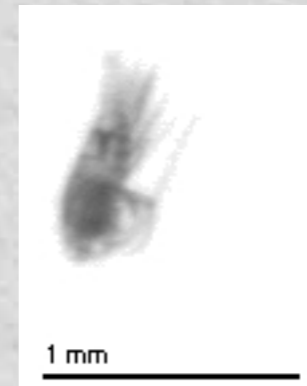
Actual machine learning workflow

OBJECT
id
lat/lon
date
class
who
feature 1
...
feature n
cnn 1
...
cnn 50
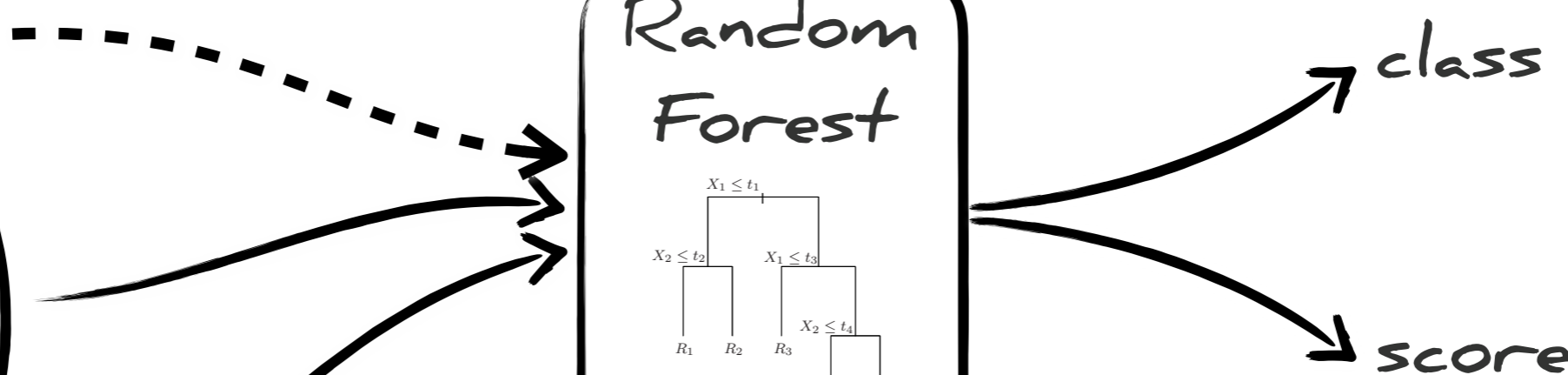
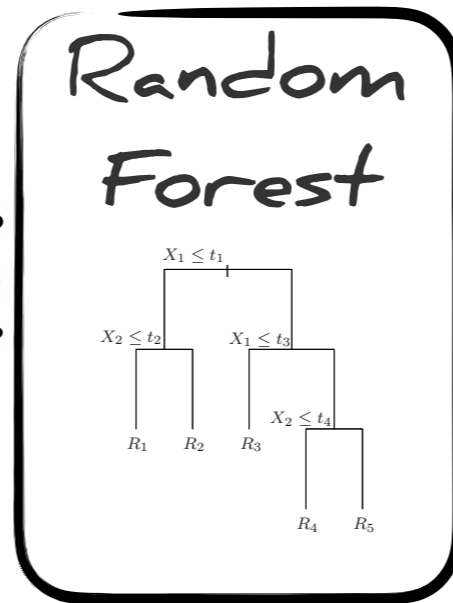| Dataset	Grouping	Random Forest (ZP)	Random Forest (SCN)	Random Forest (ZP+SCN)	SparseConvNet
flowcam	group1	82.65%	82.98%	86.64%	69.22%
flowcam	group2	82.56%	83.05%	86.54%	68.15%
uvp5ccelter	group1	85.89%	79.96%	86.70%	52.36%
uvp5ccelter	group2	85.91%	79.74%	86.63%	51.68%
zoocam	group1	87.49%	90.30%	92.16%	88.64%
zoocam	group2	92.74%	93.46%	95.31%	89.85%
zooscan	group1	71.14%	78.90%	79.96%	61.09%
zooscan	group2	70.84%	78.22%	79.62%	60.58%

Usage statistics (as of yesterday)
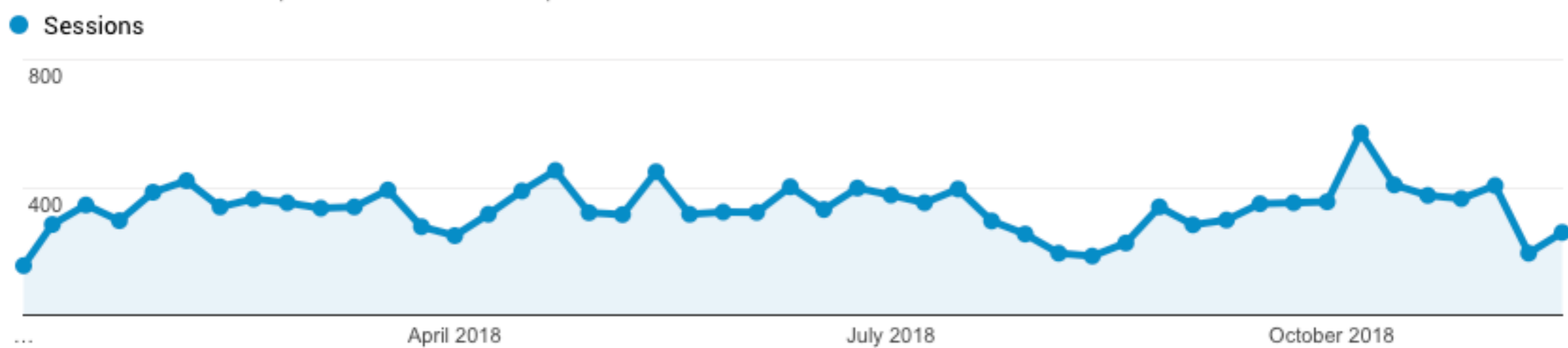
68M objects

43% validated

800 projects

504 registered users, in 146 organisations

92k taxa, 1200 actually used
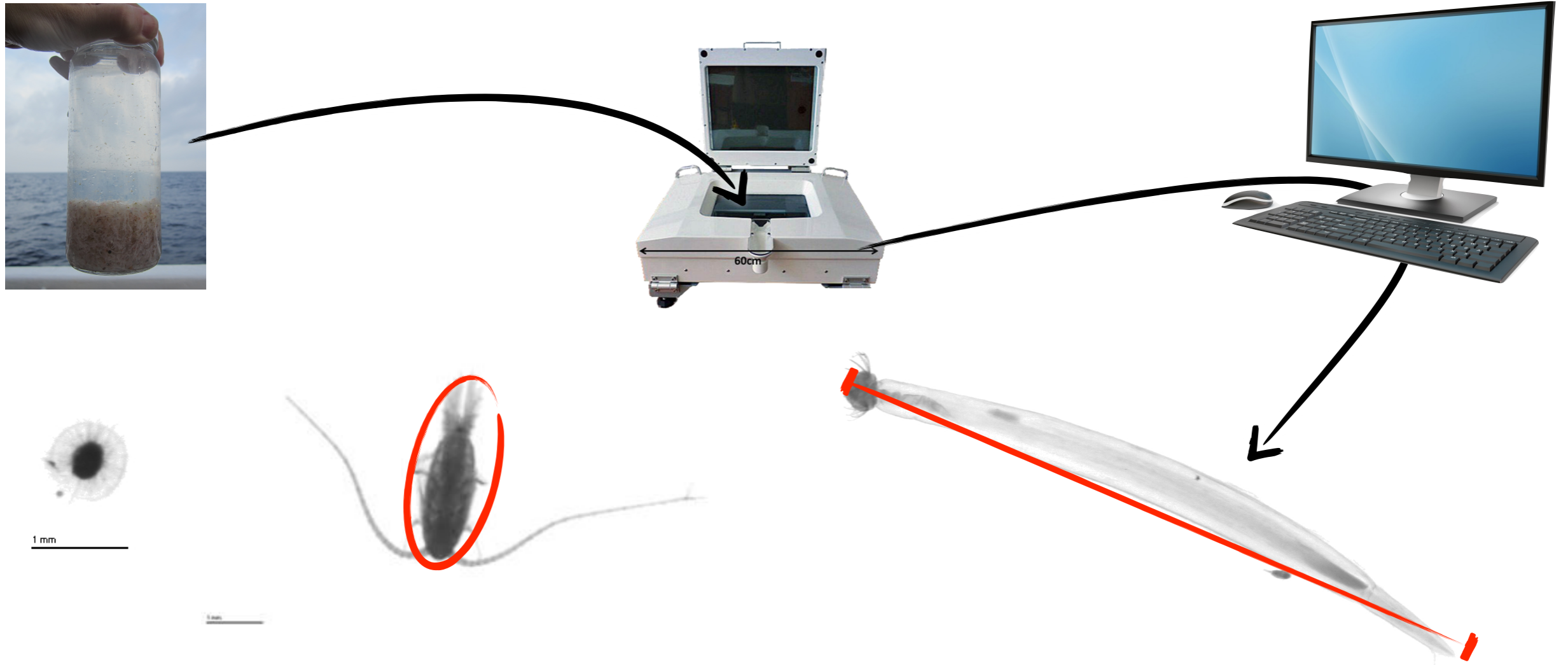
10 to 20k classifications/user/day

An objective measure of morphological diversity
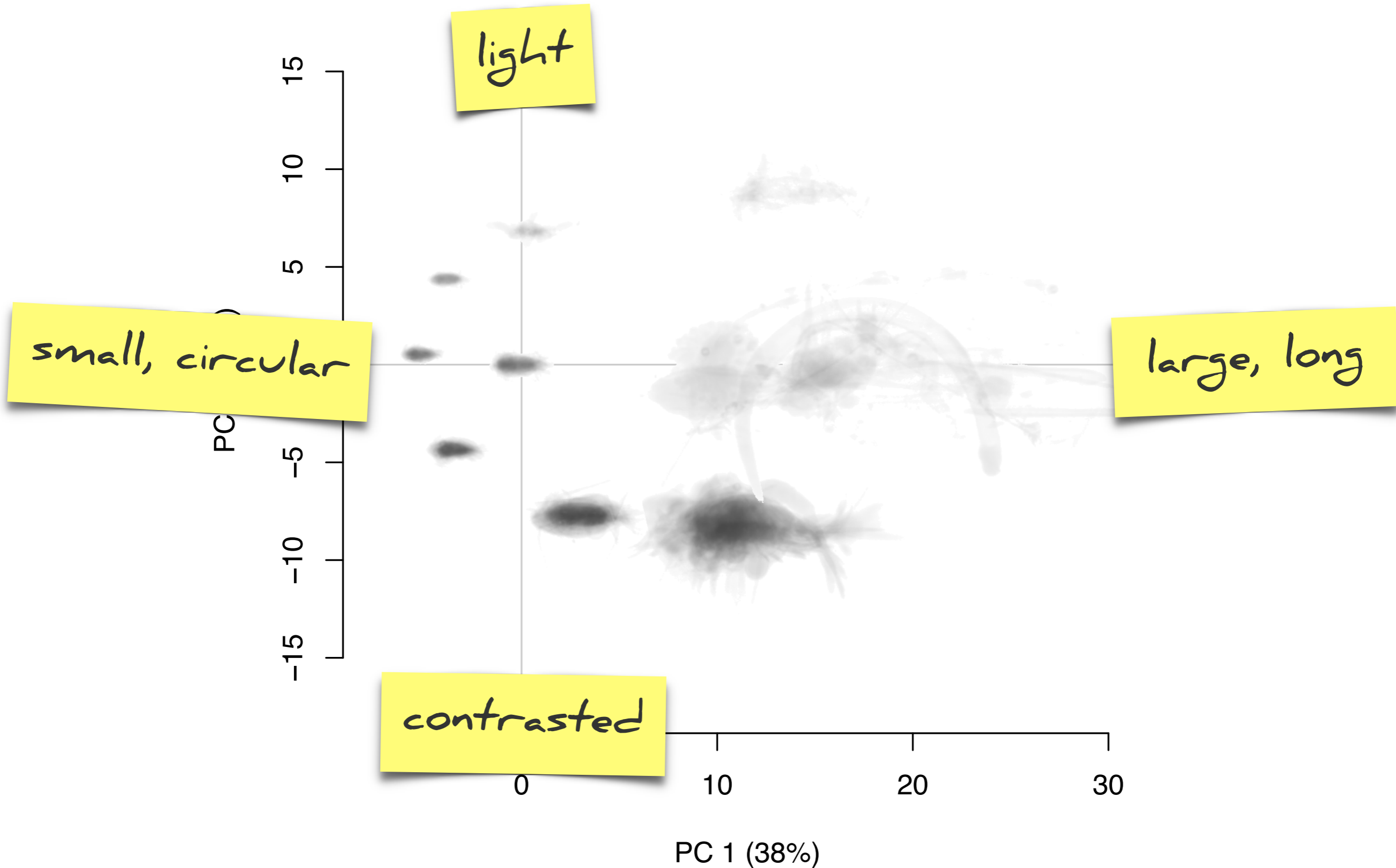
How to cut along a gradient

Automatic extraction of morphological features

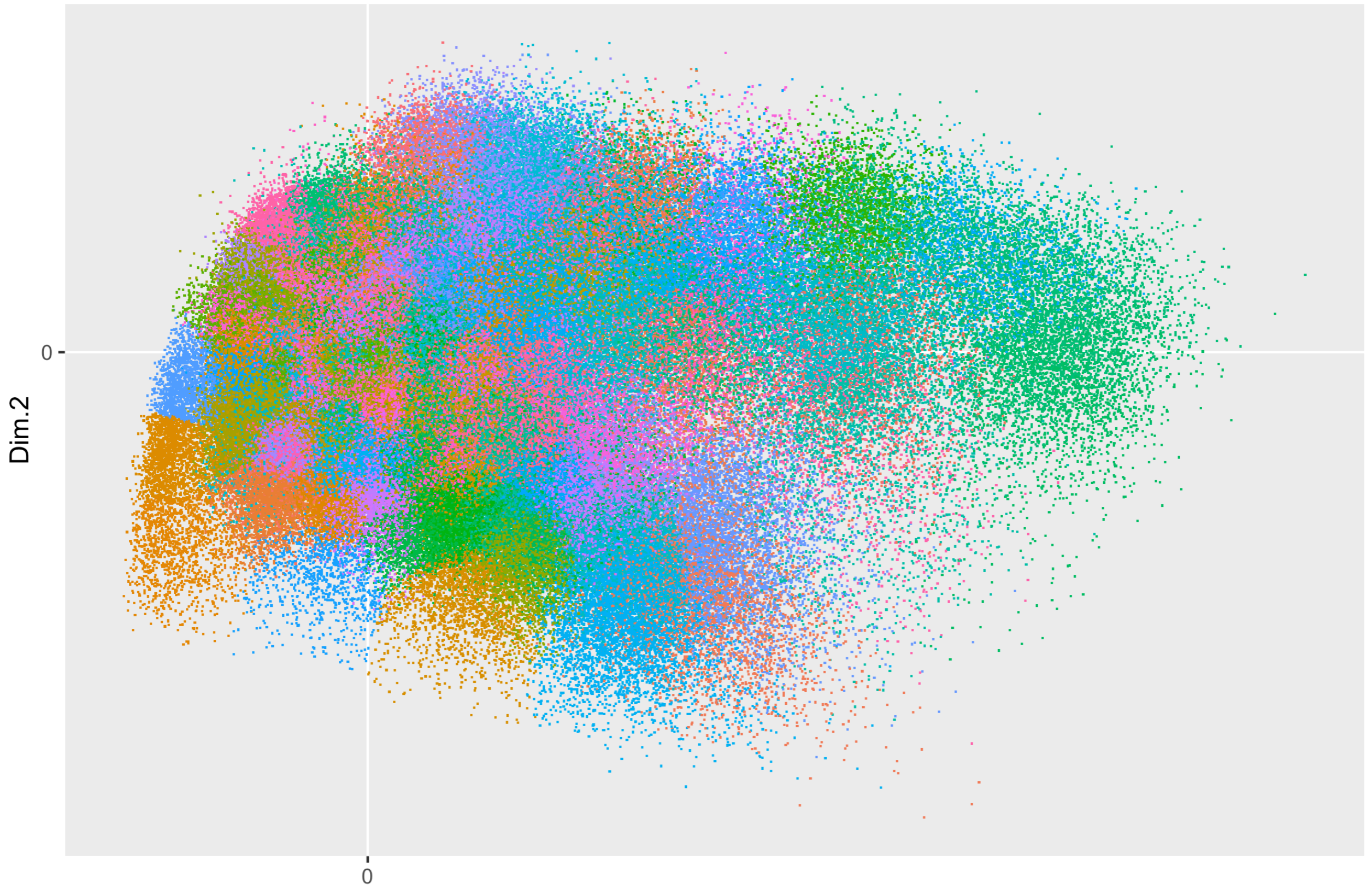|	feature 1	...	feature m
ind 1			
ind 2			
ind 3			
ind 4			

Plankton in morphological space

light

small, circular

large, long

contrasted

PC 1 (38%)
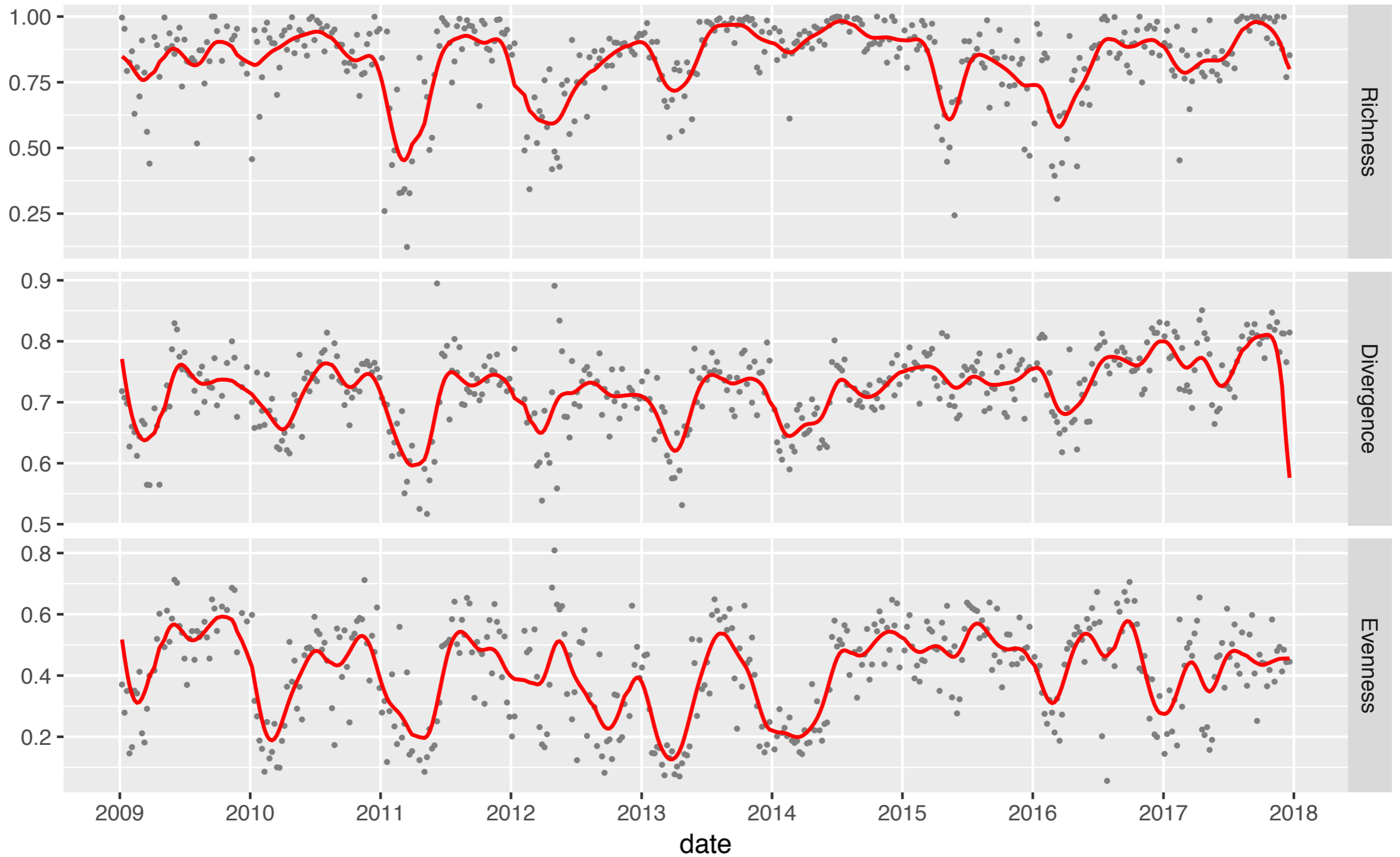
Define "morphs" (through objective clustering)

Time series of morphological diversity indices

How to do computer science as a biologist?

You have to **learn** some of it, no way around it

Various **collaborations** ongoing

Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis (I3S)

Facebook Artificial Intelligence Research (FAIR)

Google France

SAP

Now, some **proposals** in the pipelines

Merci