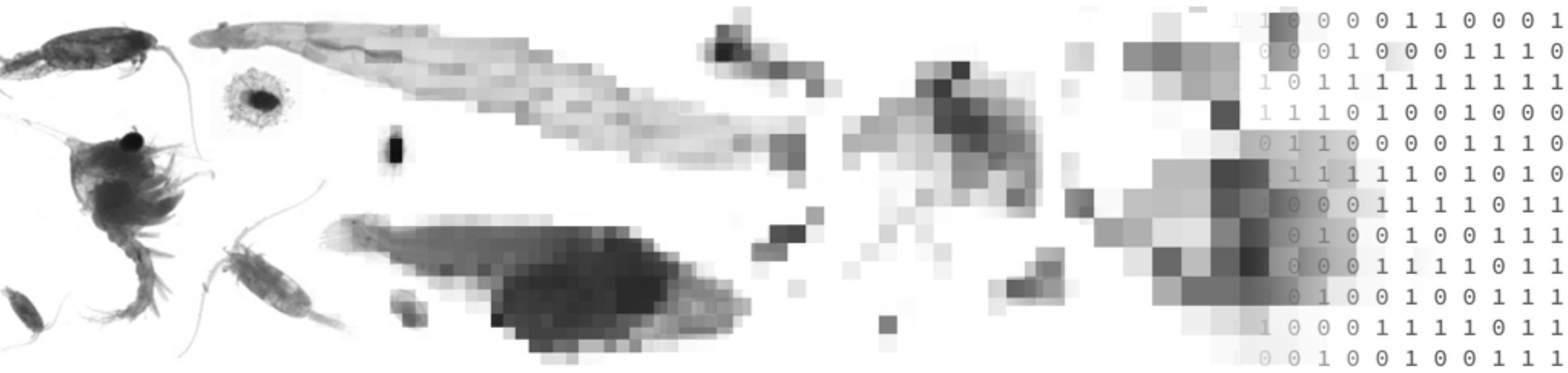
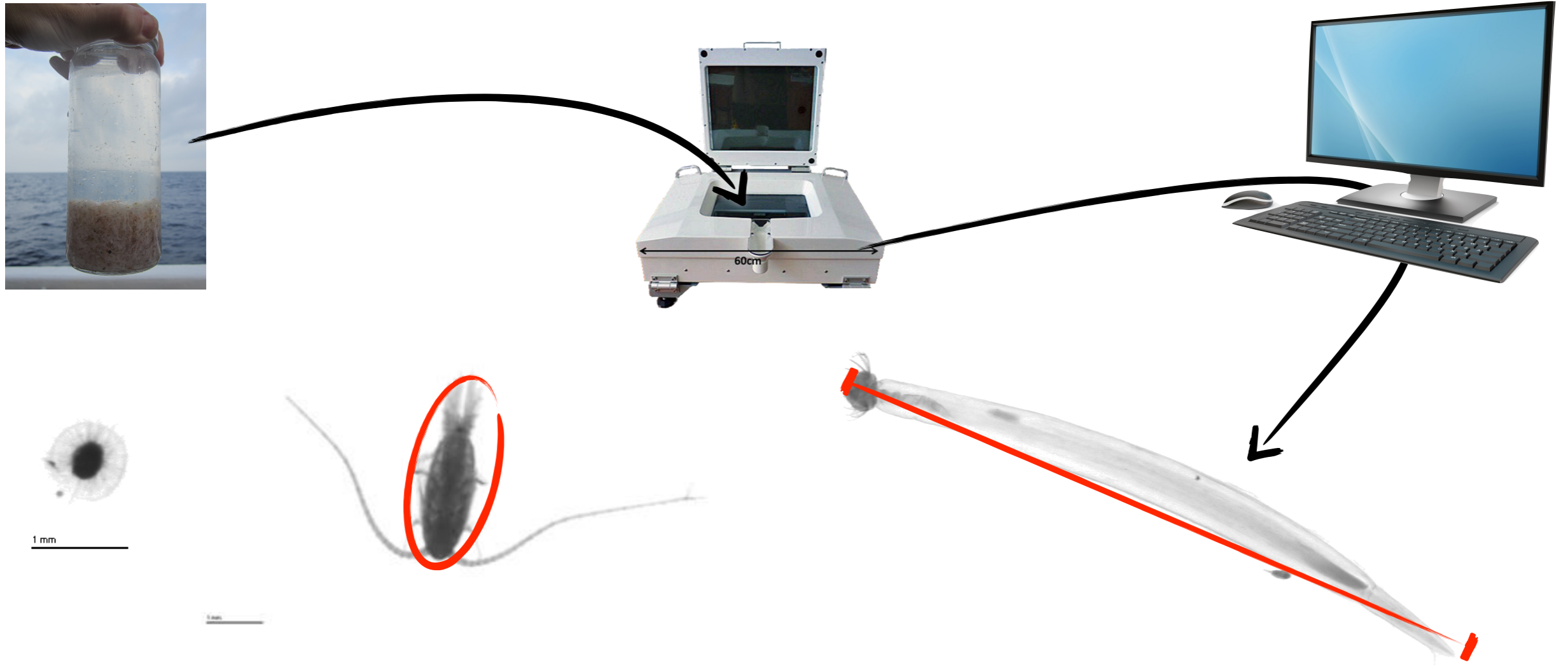


Quantitative image processing with machine learning

How to turn images into data



Automatic extraction of morphological features



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

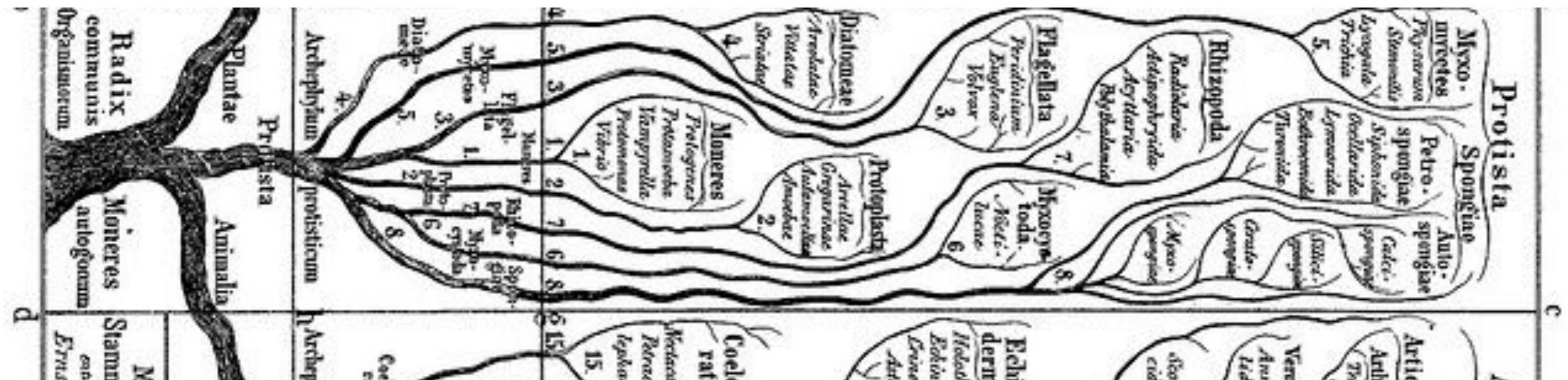


Current flow of images

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

Classification of images

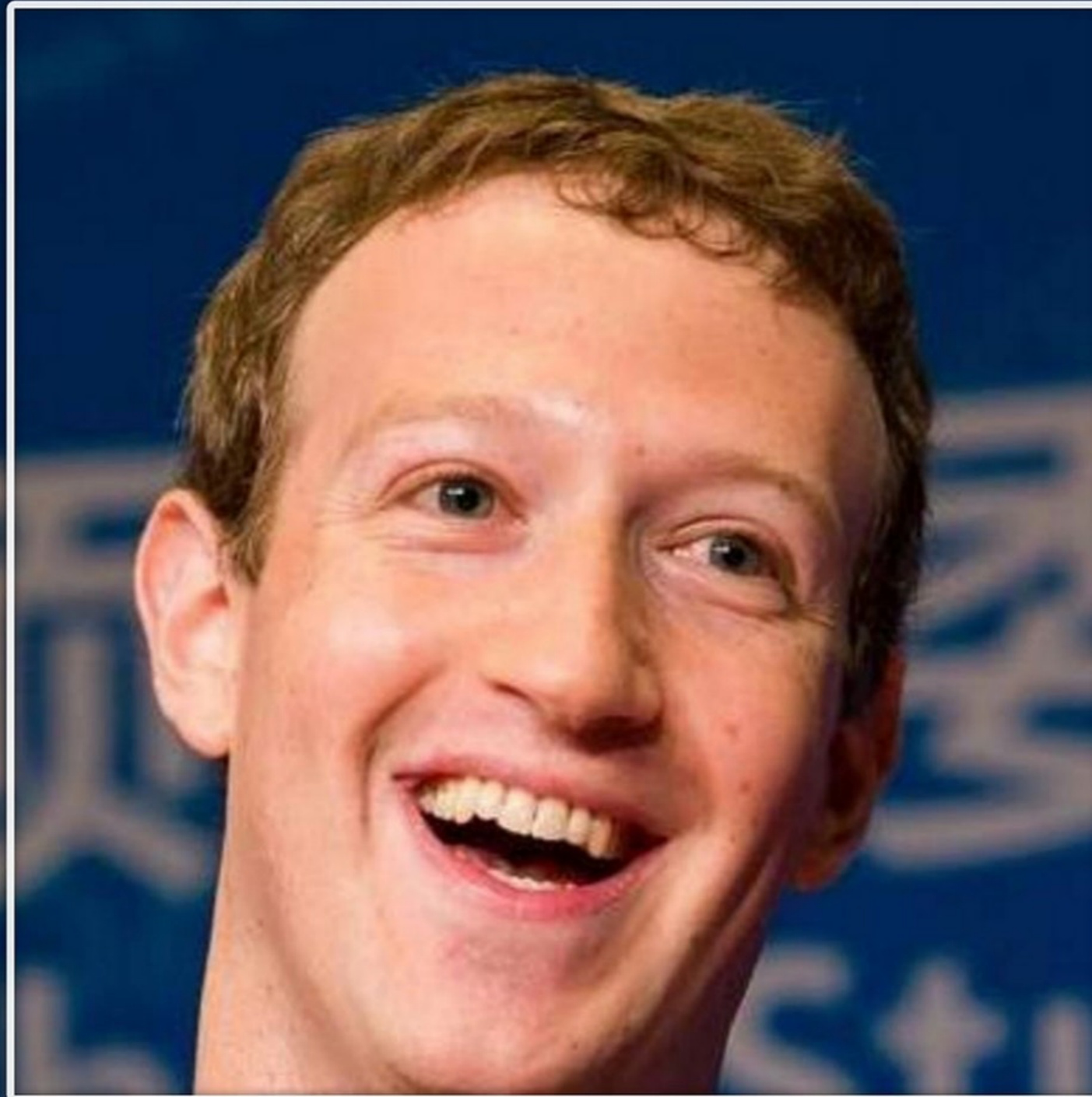
Because we like to put things in boxes



Demo

<http://ecotaxa.obs-vlfr.fr/prj/1334>

Deep Learning... how he does this:



mark zuckerberg|



Mark Zuckerberg

11,814 like this. · Business P...



A trendy topic

● Deep learning
Field of study

● Random forest
Topic

● Support vector ma...
Topic

+ Add comparison

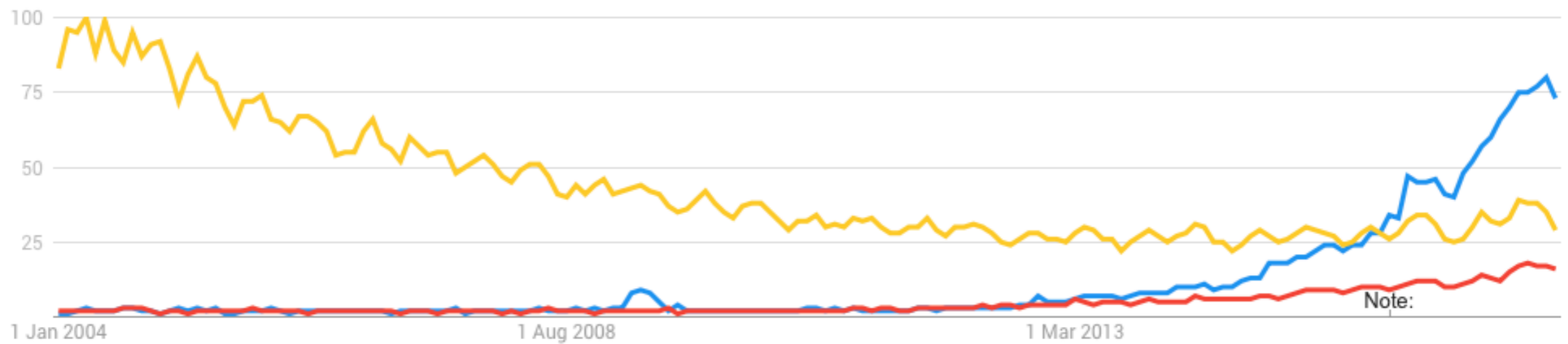
Worldwide ▾

2004 - present ▾

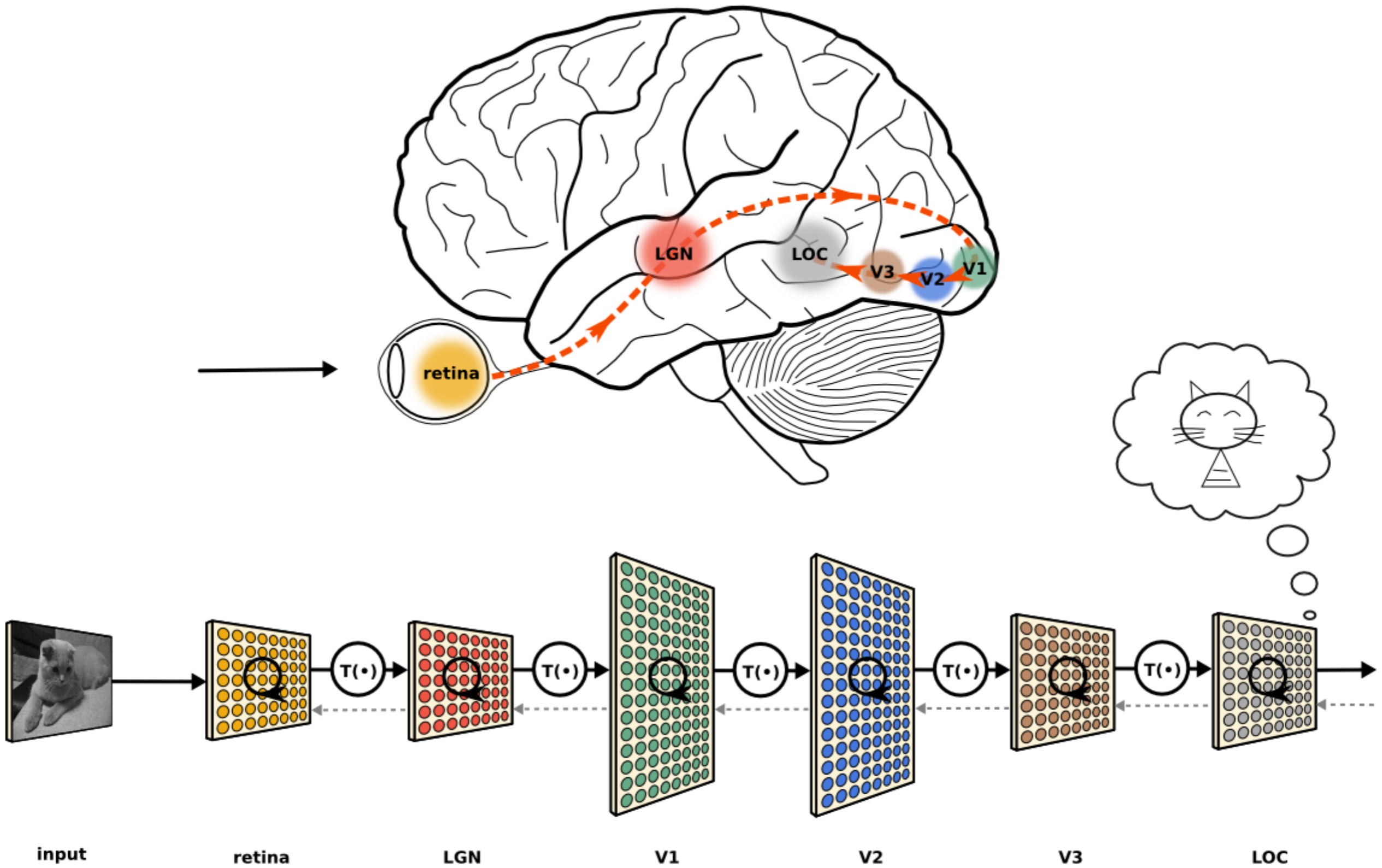
All categories ▾

Web Search ▾

Interest over time ?



Works like your brain



Inner workings of a CNN

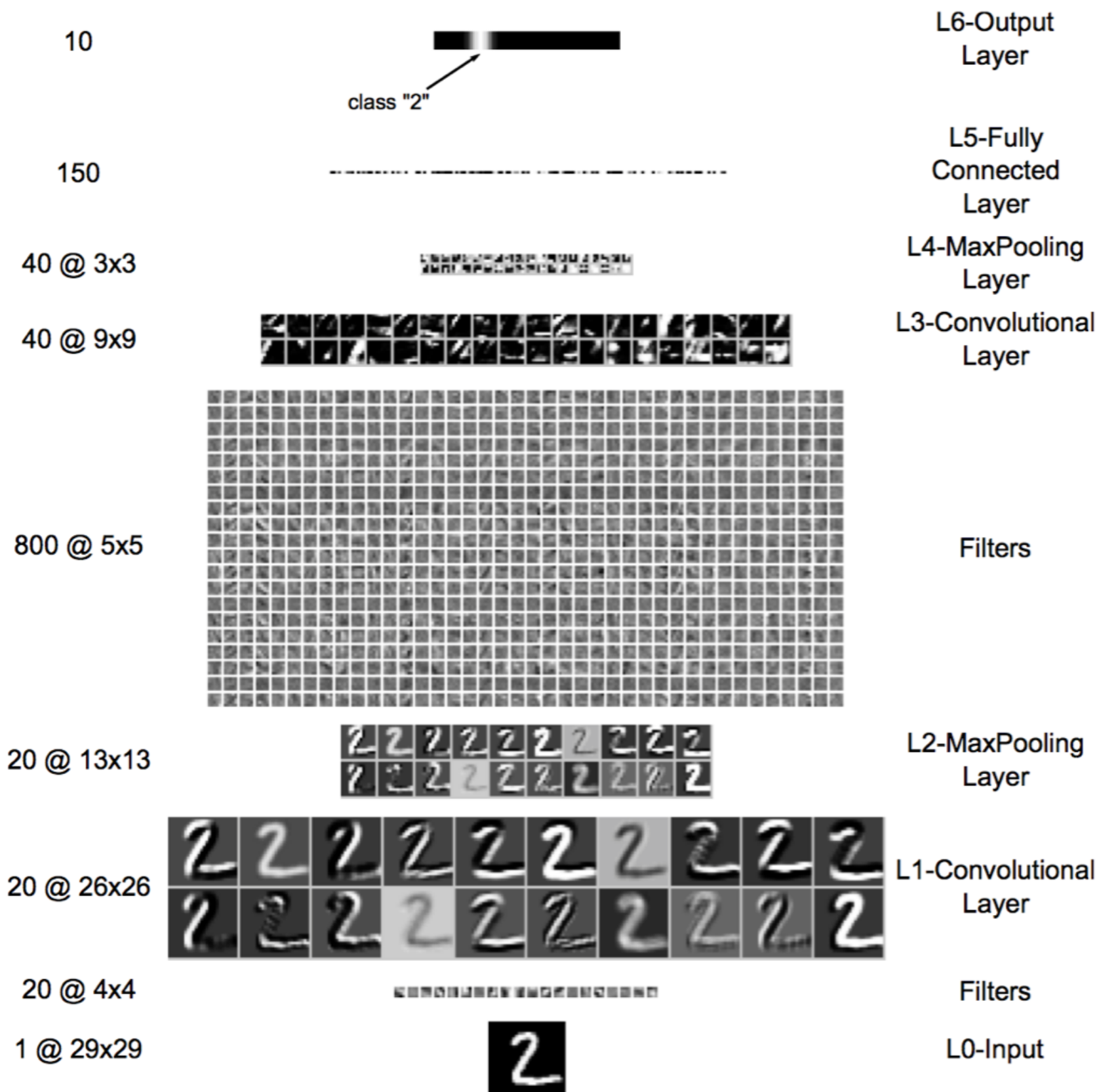
CNN = Convolutional Neural Network

It extracts “morphological” features

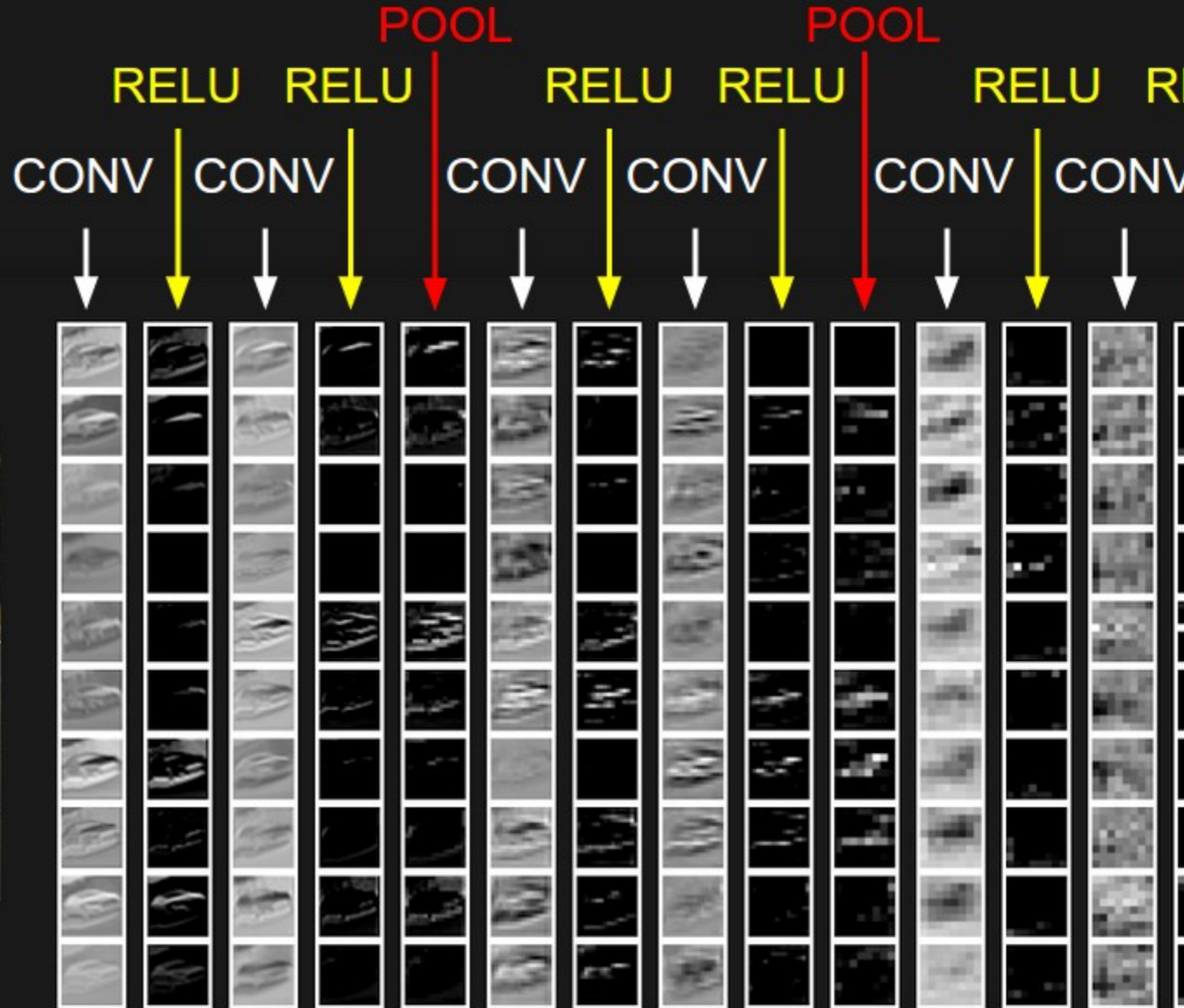
(and classifies objects)

Features are “seen” at various scales

Training the network means assigning **weights**



+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



Completed • \$175,000 • 1,049 teams
National Data Science Bowl

Mon 15 Dec 2014 – Mon 16 Mar 2015 (22 months ago)

Dashboard

Home

Data
Make a submission

Information

Description
Evaluation
Rules
Prizes
About the NDSB
Timeline
Tutorial

Forum

Leaderboard

Public
Private

My Team

GitHub

My Submissions

Private Leaderboard

1. Deep Sea
2. Happy Lantern Festival
3. Poisson Process
4. Junonia
5. Deepsea Challenger
6. AuroraXie
7. Maxim Milakov
8. Ilya Kostrikov
9. old-ufo
10. nagadomi

Forum (154 topics)

scikit-learn Random Forest memory problem
3 months ago

Install Theano on Windows 8.1 with GPU enabled: pycuda installation problems
4 months ago

caffe training curves
5 months ago

Does anyone use caffe? How could I produce a test result?
9 months ago

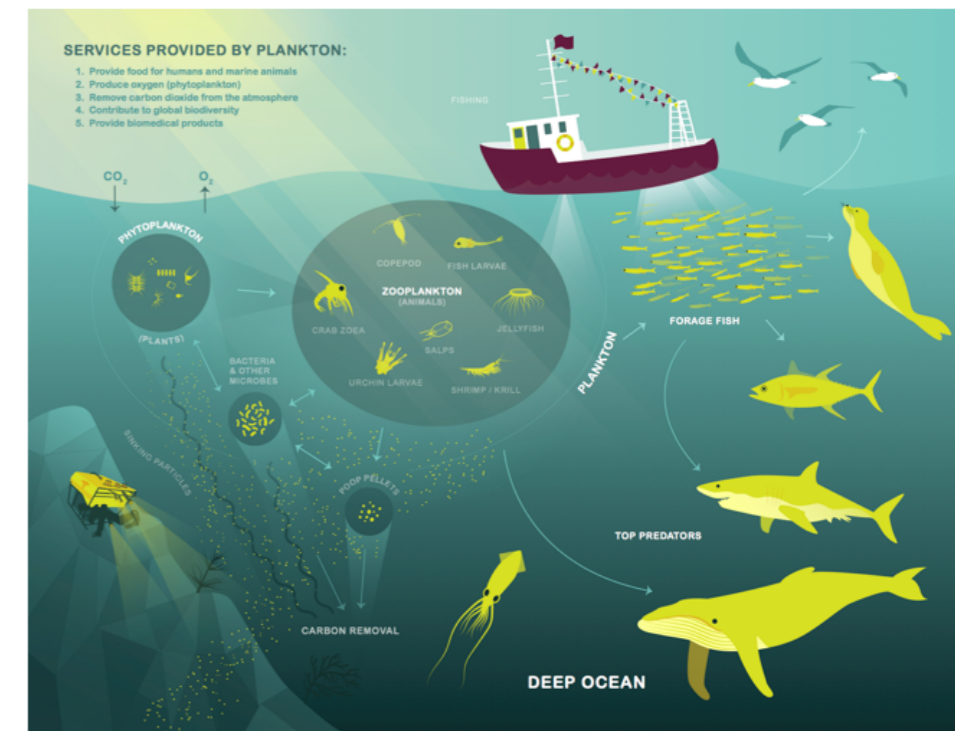
Caffe? How to generate the prediction from caffe output?
10 months ago

Can someone explain what batch size is doing in convolutional NNs?
13 months ago

Competition Details » [Get the Data](#) » [Make a submission](#)

Predict ocean health, one plankton at a time

Plankton are critically important to our ecosystem, accounting for more than half the primary productivity on earth and nearly half the total carbon fixed in the global carbon cycle. They form the foundation of aquatic food webs including those of large, important fisheries. Loss of plankton populations could result in ecological upheaval as well as negative societal impacts, particularly in indigenous cultures and the developing world. Plankton's global significance makes their population levels an ideal measure of the health of the world's oceans and ecosystems.



Traditional methods for measuring and monitoring plankton populations are time consuming and cannot scale to the granularity or scope necessary for large-scale studies. Improved approaches are needed. One such approach is through the use of an underwater imagery sensor. This towed, underwater camera system captures microscopic, high-resolution images over large study areas. The images can then be analyzed to assess species populations and distributions.

Manual analysis of the imagery is infeasible – it would take a year or more to manually analyze the imagery volume captured in a single day. Automated image classification using machine learning tools is an alternative to the manual approach. Analytics will allow analysis at speeds and scales previously thought impossible. The automated system will have broad applications for assessment of ocean and ecosystem health.

The National Data Science Bowl challenges you to build an algorithm to automate the

Why SparseConvNet?

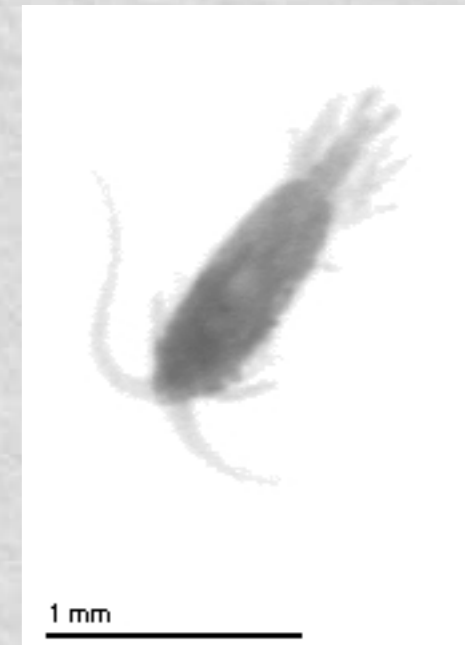
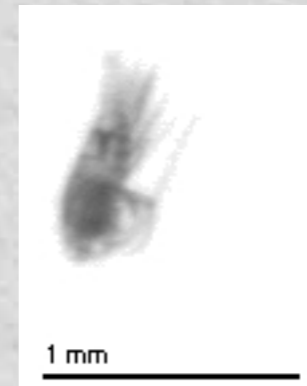
Sparsity

faster, more efficient

varying input size

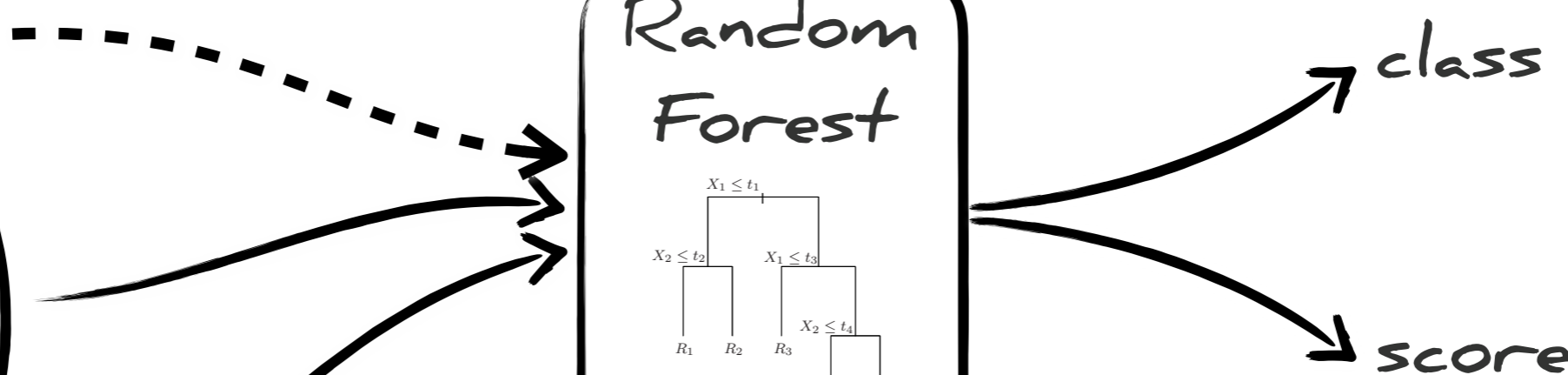
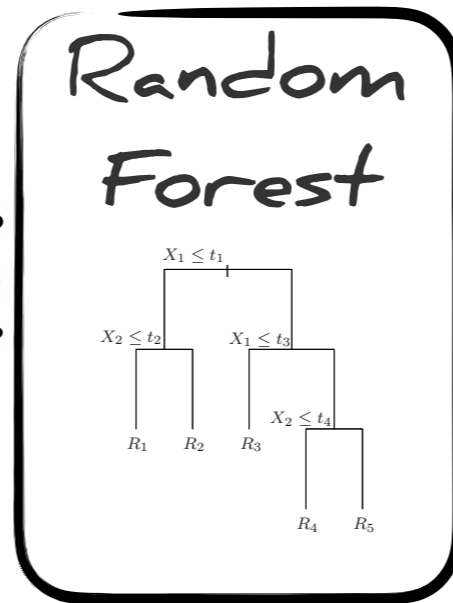
Factional Max Pooling

<https://github.com/facebookresearch/SparseConvNet>



Actual machine learning workflow

OBJECT
id
lat/lon
date
class
who
feature 1
...
feature n



<i>cnn 1</i>		Random Forest (ZP)	Random Forest (SCN)	Random Forest (ZP+SCN)	SparseConvNet
Dataset	Grouping				
flowcam	group1	82.65%	82.98%	86.64%	69.22%
flowcam	group2	82.56%	83.05%	86.54%	68.15%
<i>cnn 50</i> 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

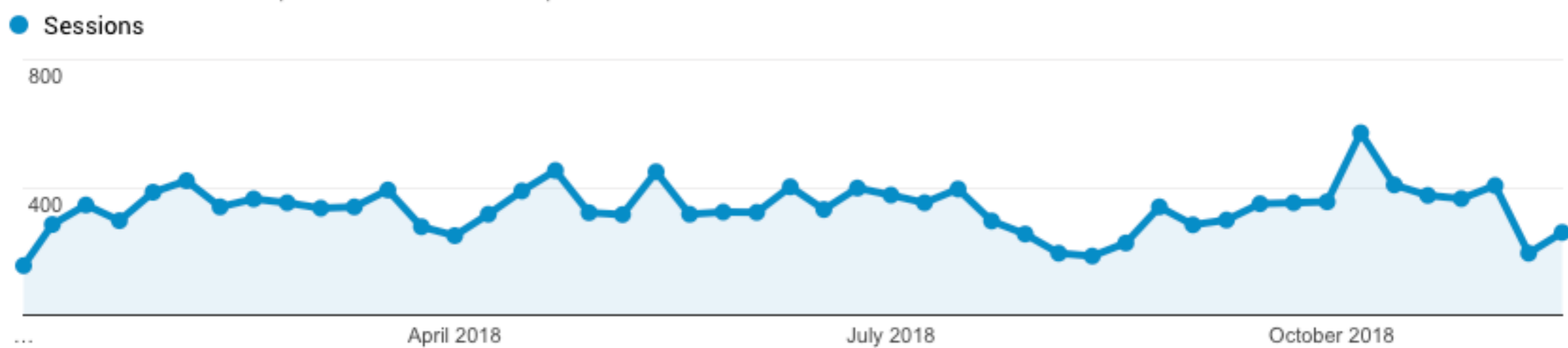
43% validated

800 projects

504 registered users, in 146 organisations

92k taxa, 1200 actually used

10 to 20k classifications/user/day

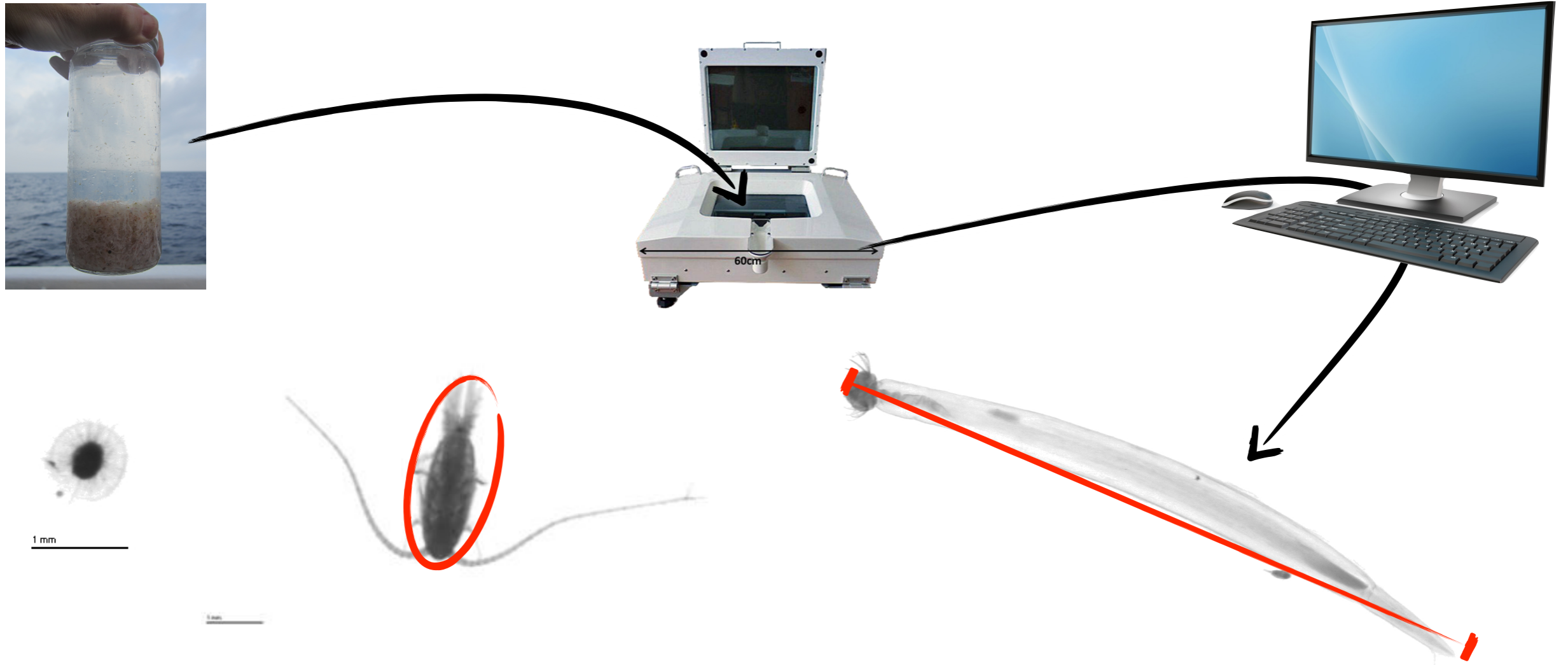


An objective measure of morphological diversity

How to cut along a gradient

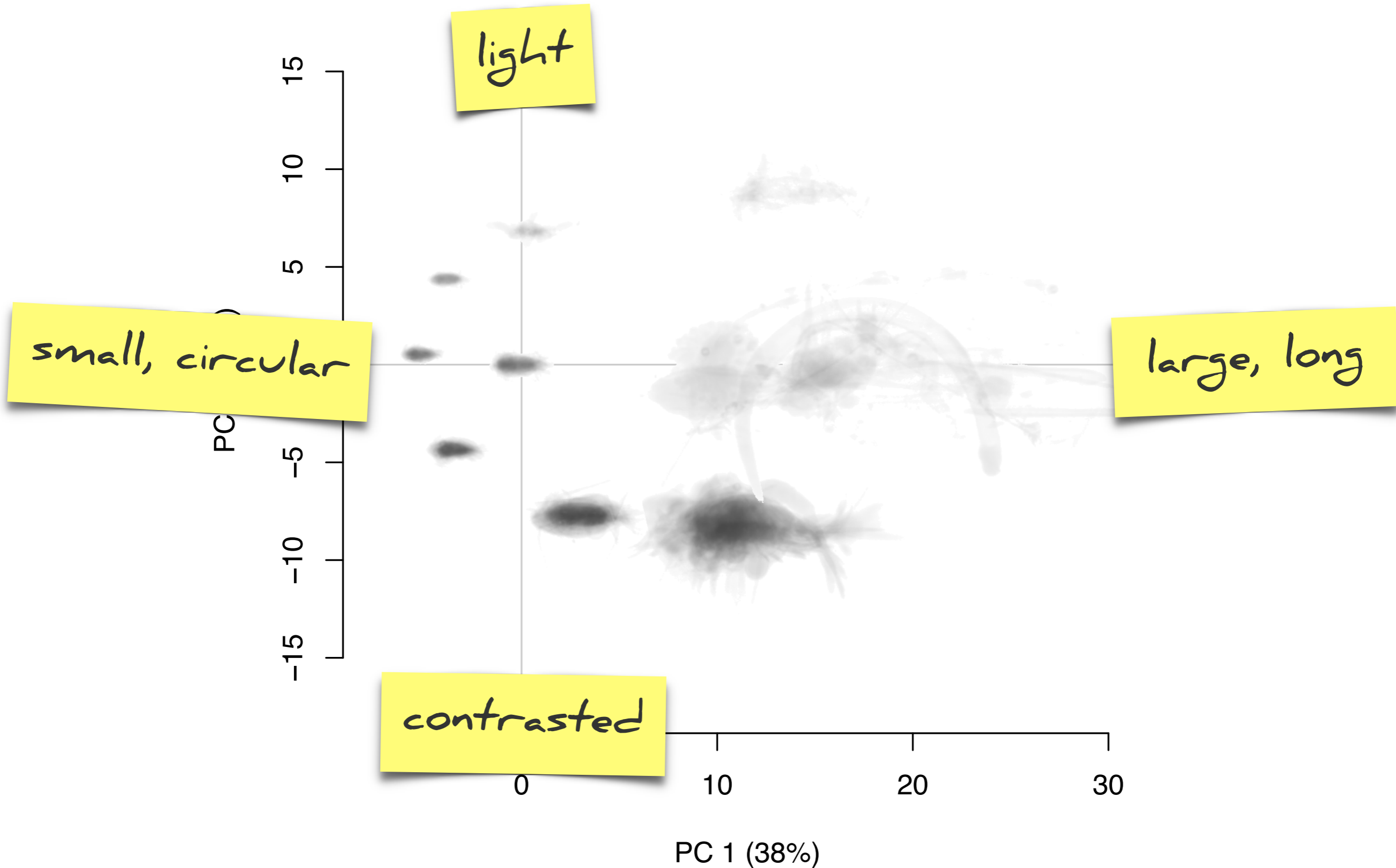


Automatic extraction of morphological features

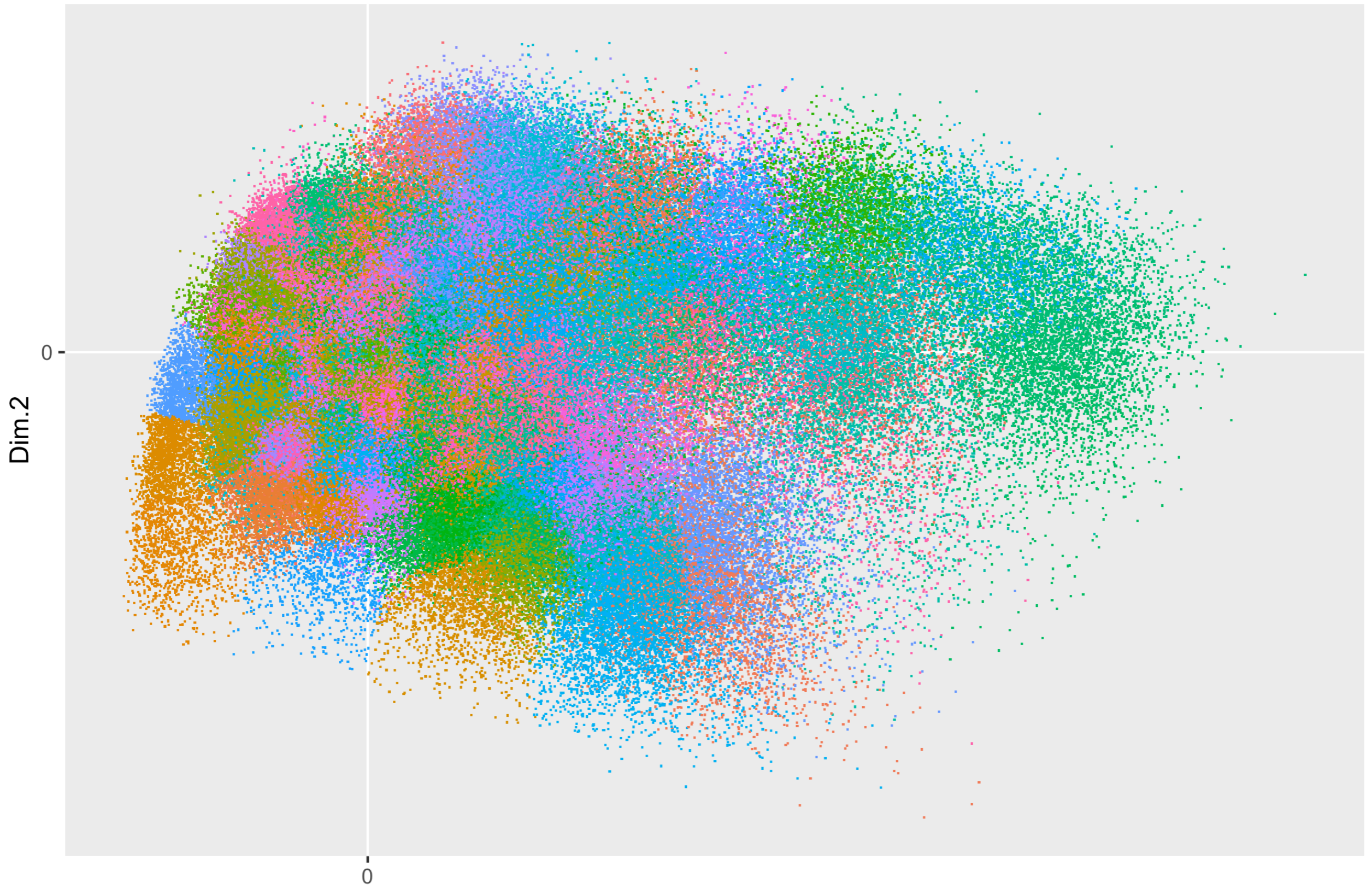


	feature 1	...	feature m
ind 1			
ind 2			
ind 3			
ind 4			

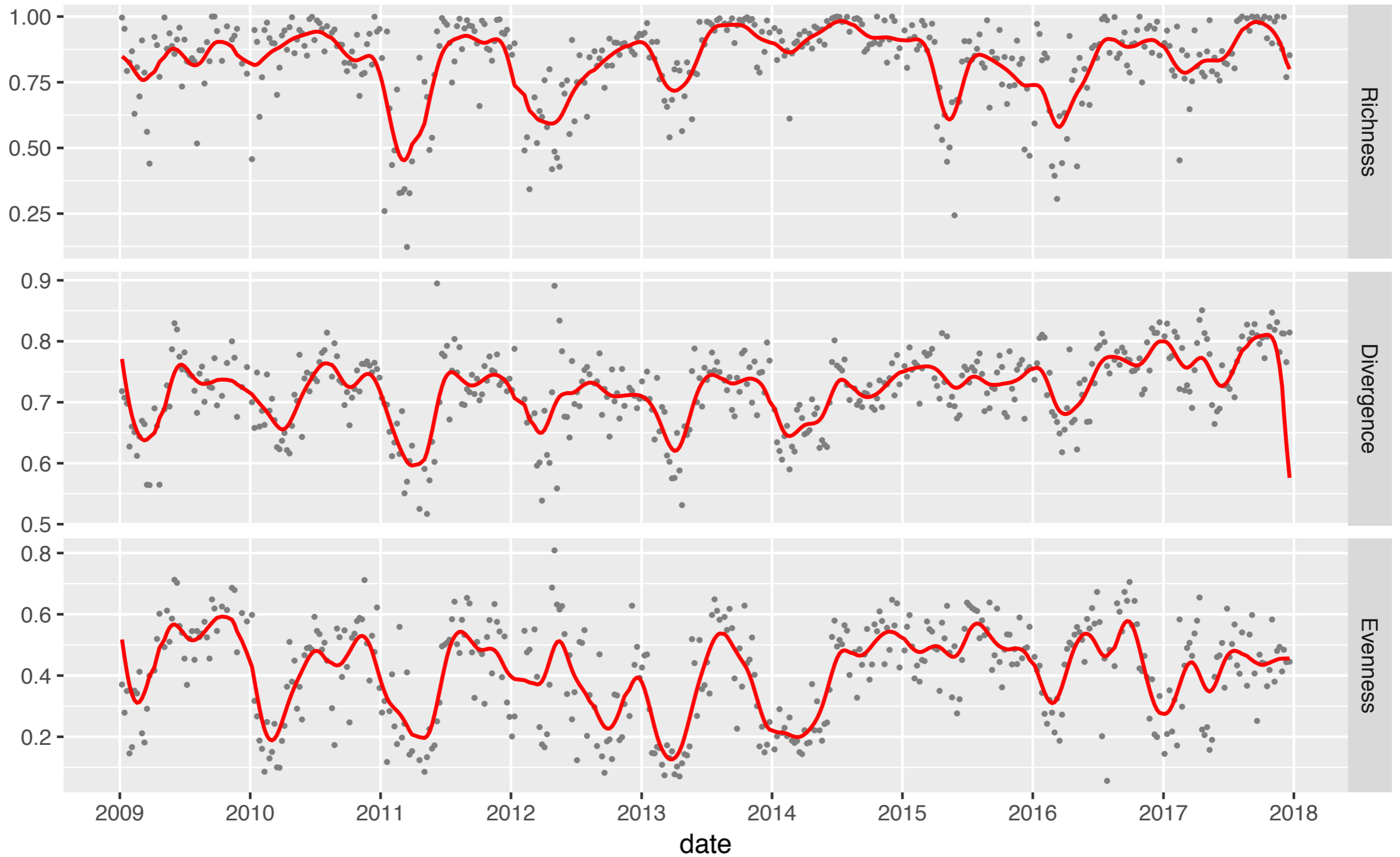
Plankton in morphological space



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



facebook

Google

SAP®



HORIZON 2020



Merci