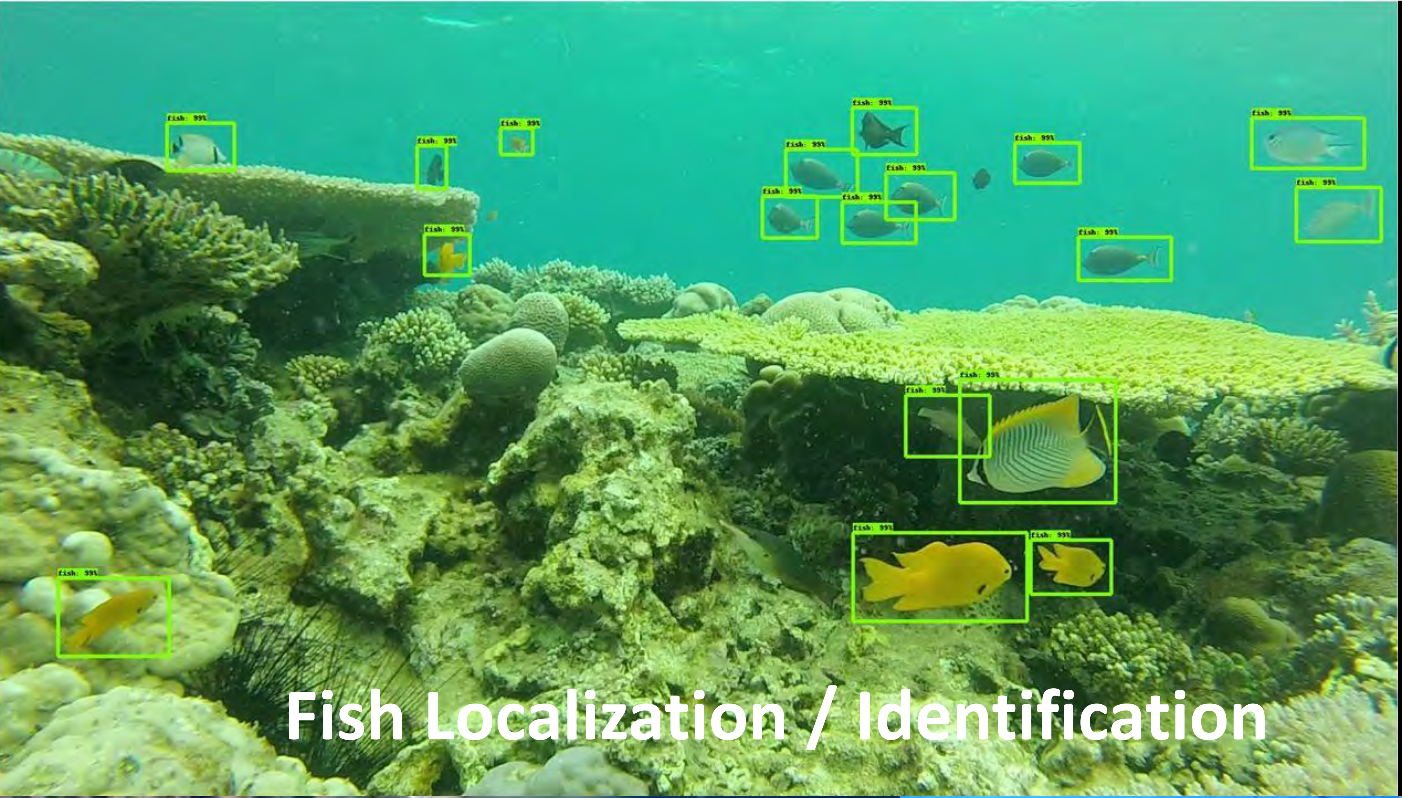
Deep-Learning for the observation of marine/terrestrial life

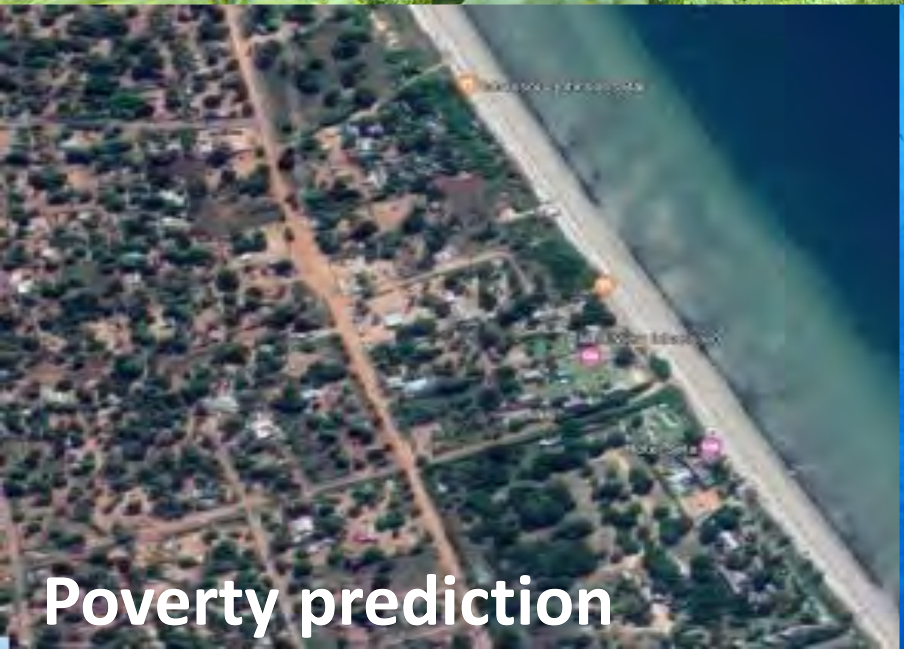
Marc Chaumont

LIRMM

Icar


Fish Localization / Identification

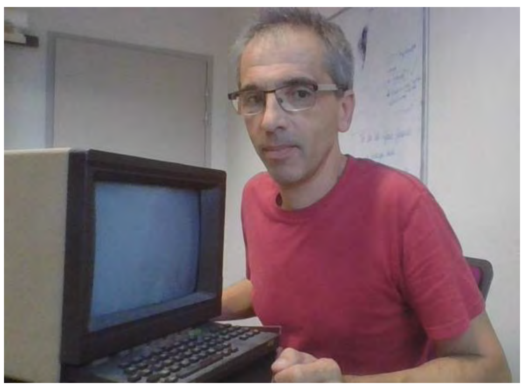

Eels
detection


Poverty prediction


Shark
localization
under DCP
with multiple
cameras
http://www.peche.pf


Species
classifications

LIRMM researchers involved



Gérard Subsol

Working on several applications of 2D and 3D image processing.

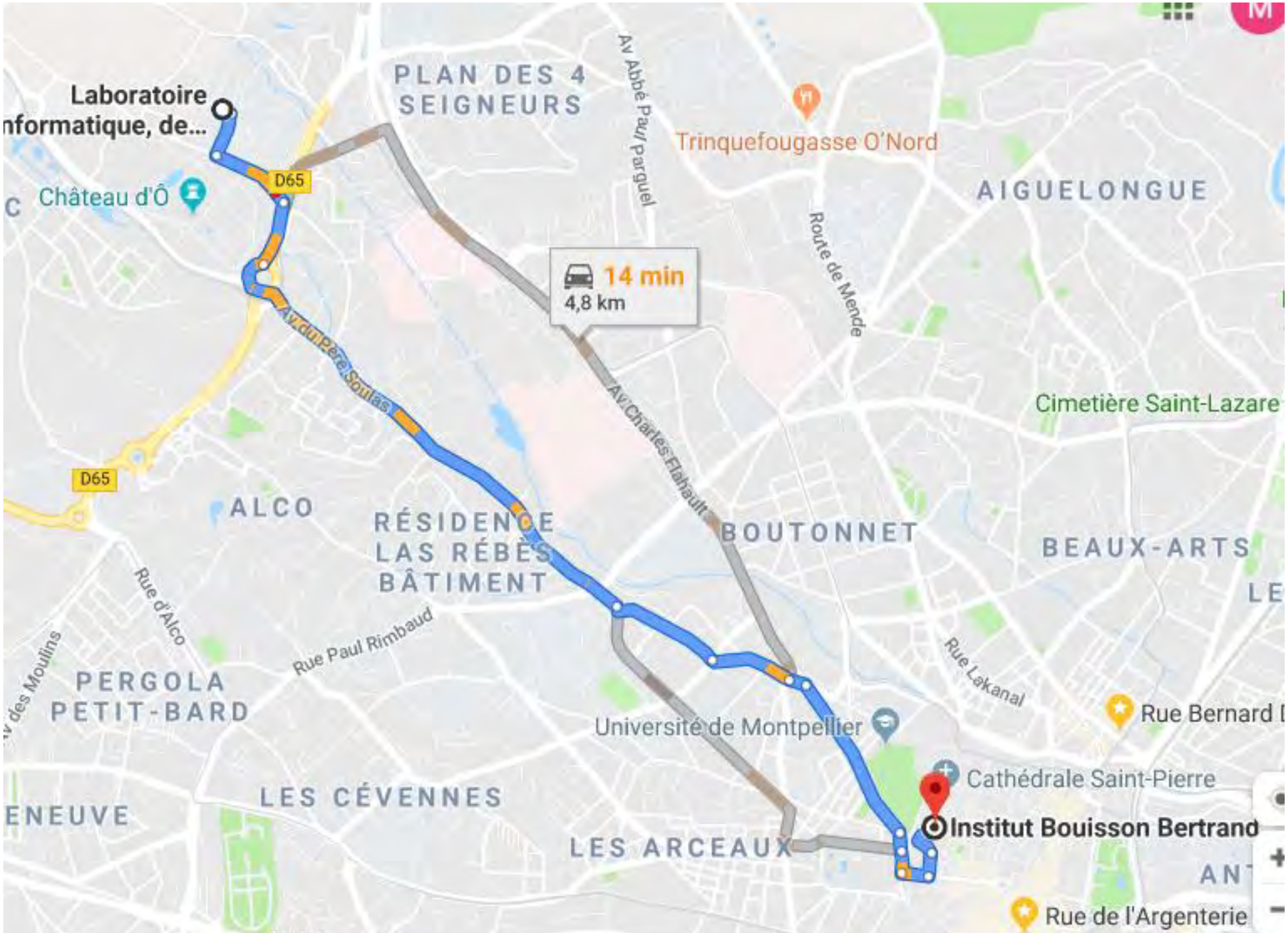
Computer science, image processing, modelling

Marc Chaumont

Steganography and more generally image and video processing

Image and video processing, Deep Learning, Data-Mining.

LIRMM (14 min from here)



Buildings

LIRMM

Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier

Historical building

New building

Outline

- Few words on Deep Learning

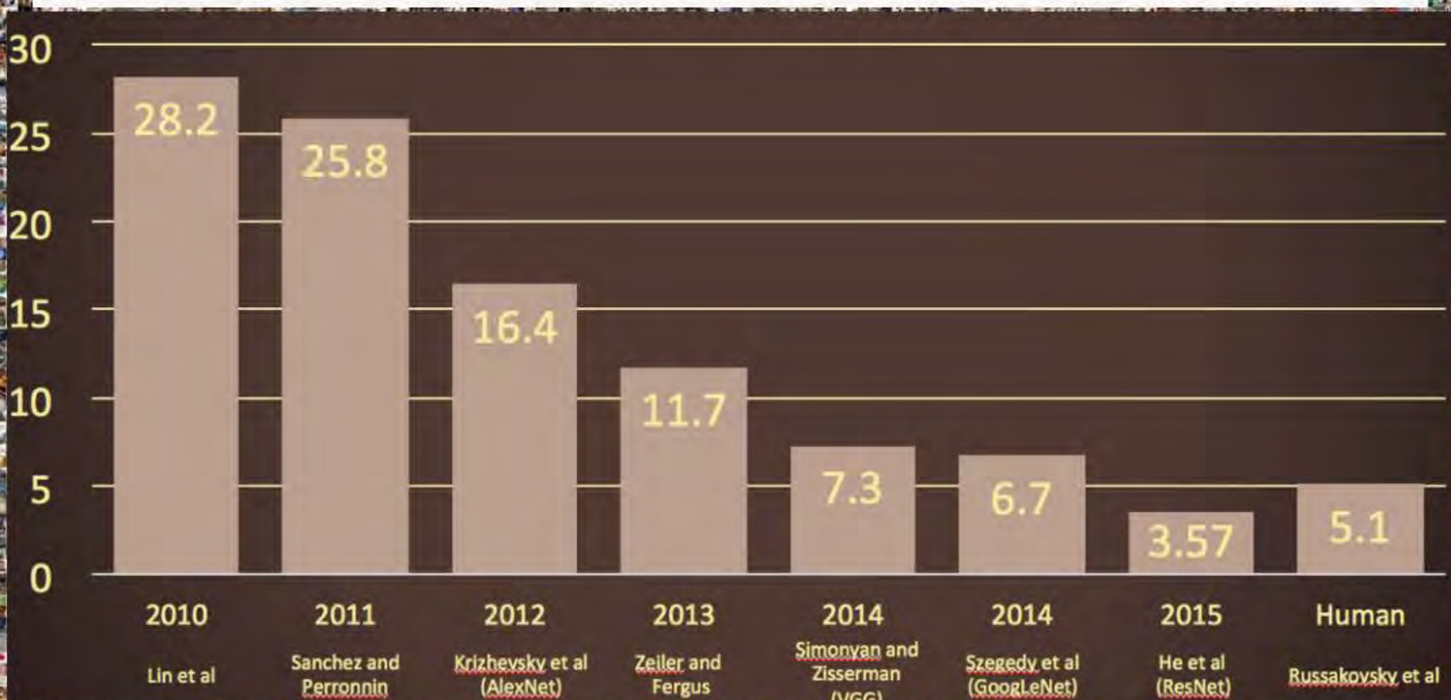
- Few projects done in our « subset »-team (Gérard and me)

IMAGENET Large Scale Visual Recognition Challenge

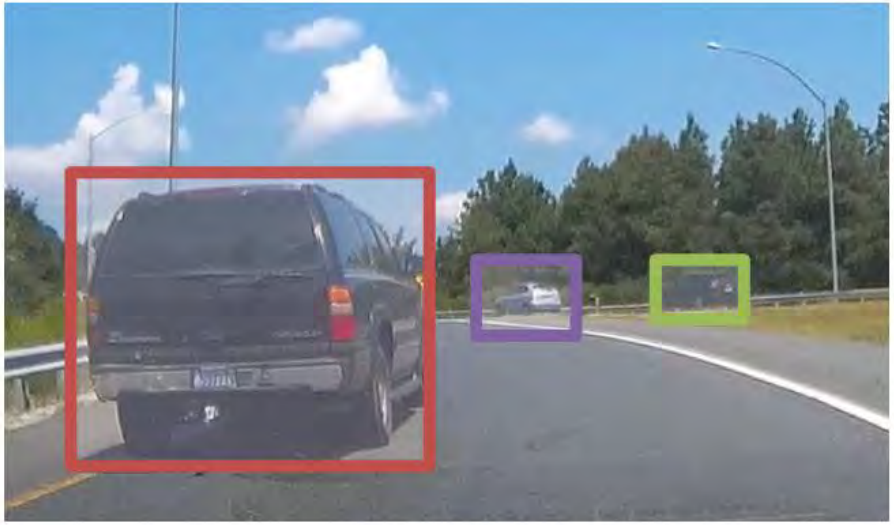
The Image Classification Challenge:

1,000 object classes

1,431,167 images

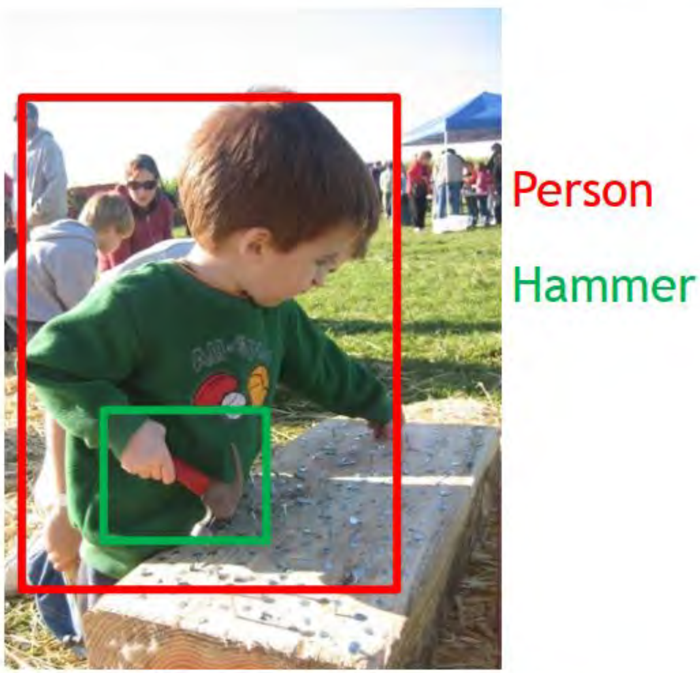


Russakovsky et al. arXiv, 2014

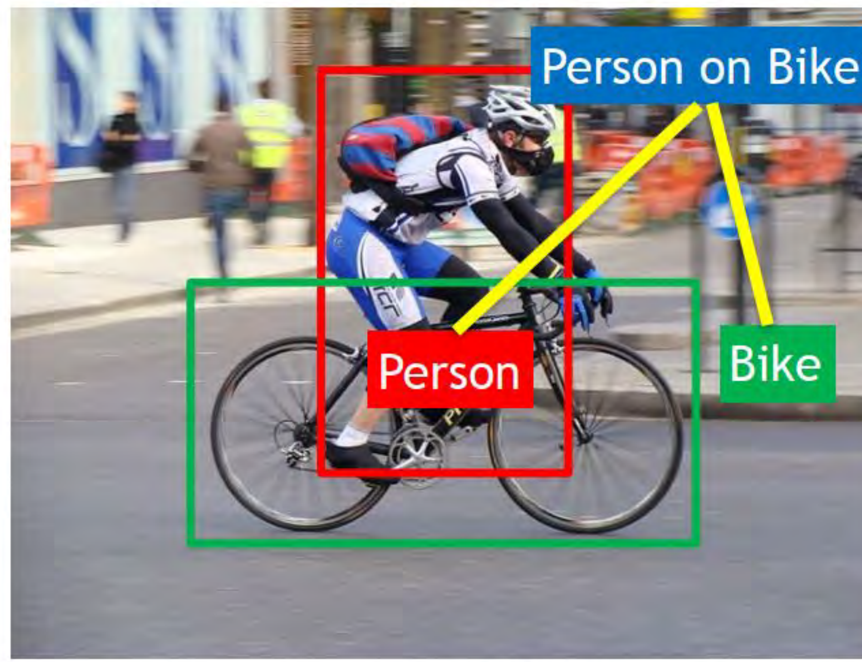


This image is licensed under CC BY-NC-SA 2.0; changes made

- Object detection
- Action classification
- Image captioning
- ...



This image is licensed under CC BY-SA 2.0; changes made

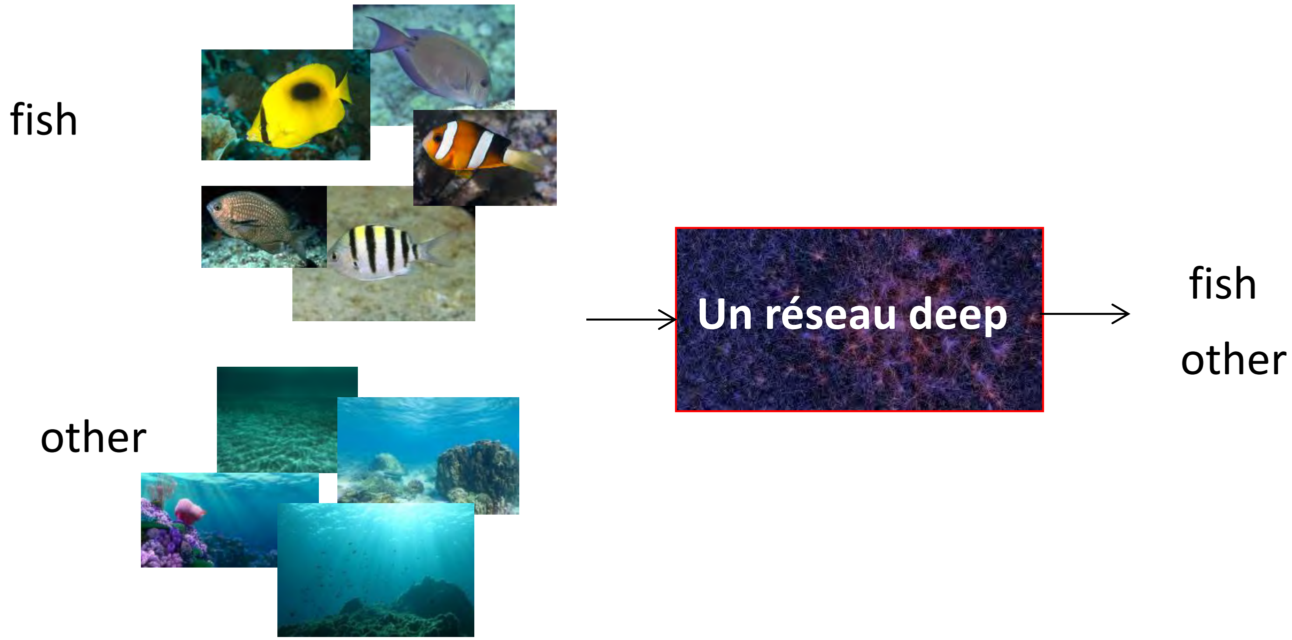


This image is licensed under CC BY-SA 3.0; changes made

The learning protocol (supervised case)

- STEP 1) We « show » to the « network » exemples et counter-exemples

THE LEARNING

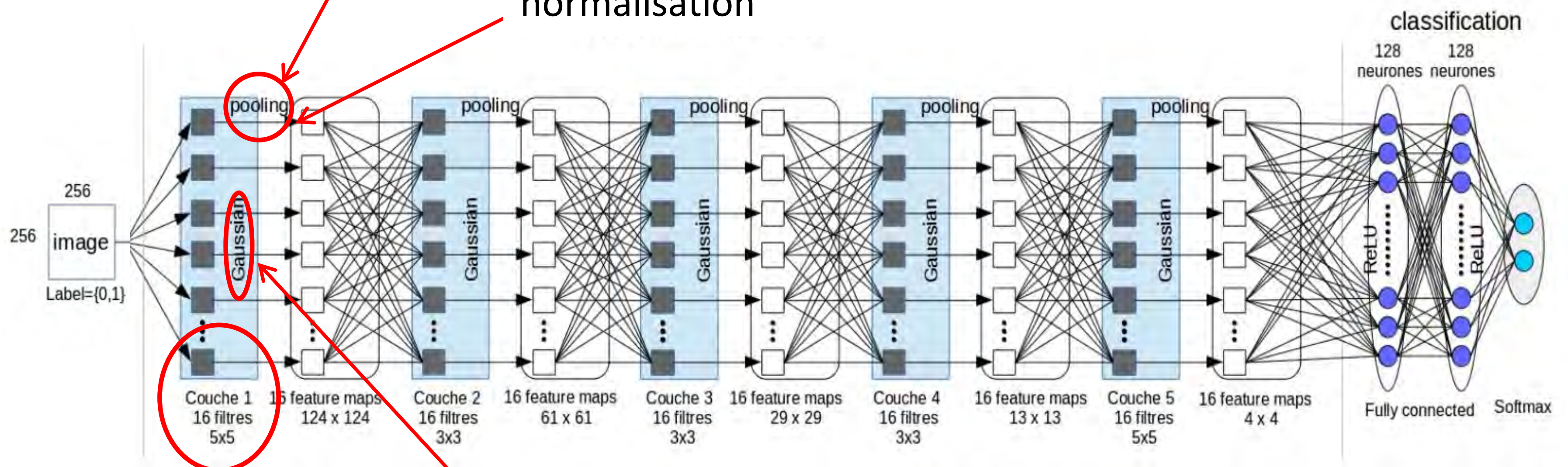


- STEP 2) We use the network ☺

A CNN
Convolutional Neural Network

Average or max [+ sub-sampling]

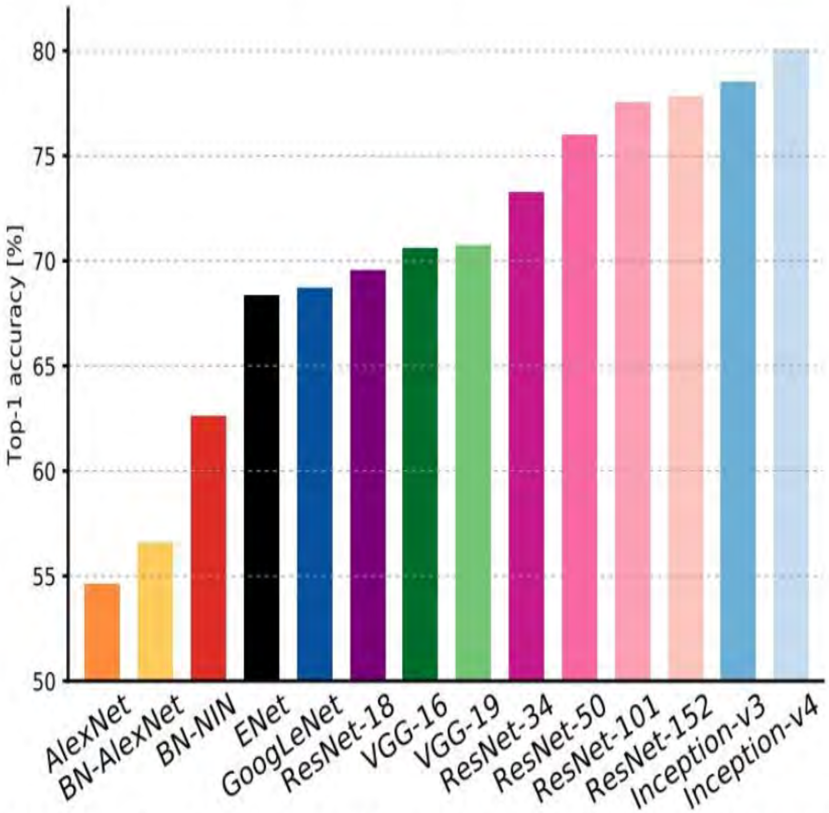
normalisation

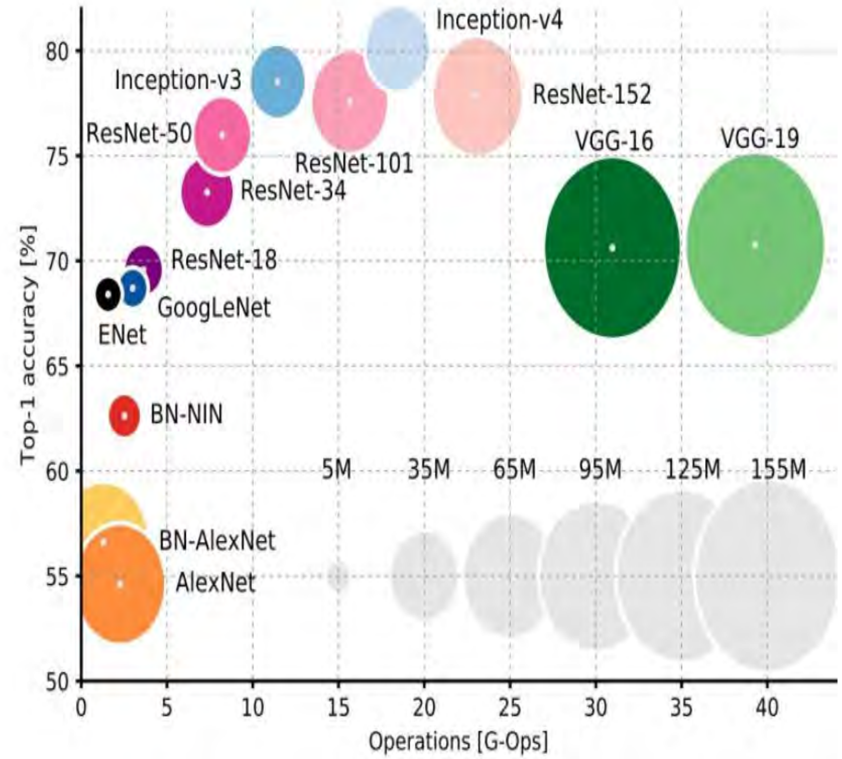


convolutions

Non linear function
(= activation function)

Comparing complexity...



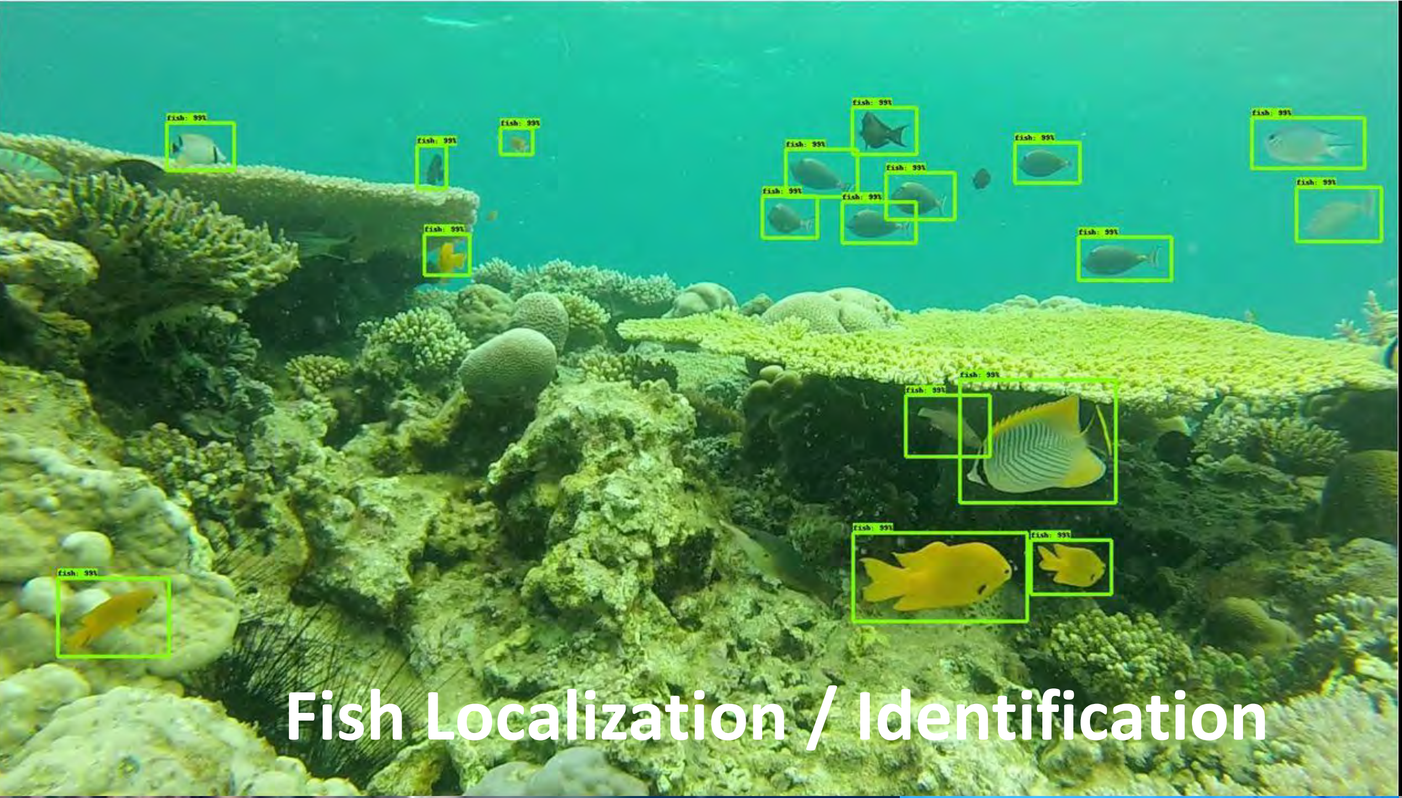


An Analysis of Deep Neural Network Models for Practical Applications, 2017.

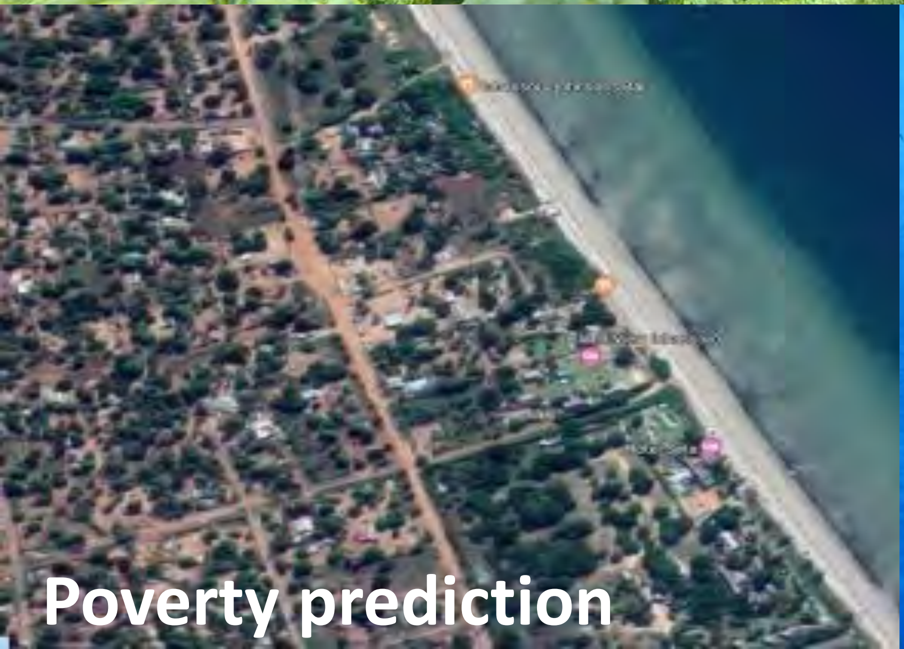
Outline

- Few words on Deep Learning

- Few projects done in our « subset »-team (Gérard and me)


Fish Localization / Identification


Eels
detection


Poverty prediction


Shark
localization
under DCP
with multiple
cameras
http://www.peche.pf


Species
classifications

Counting and identification of species

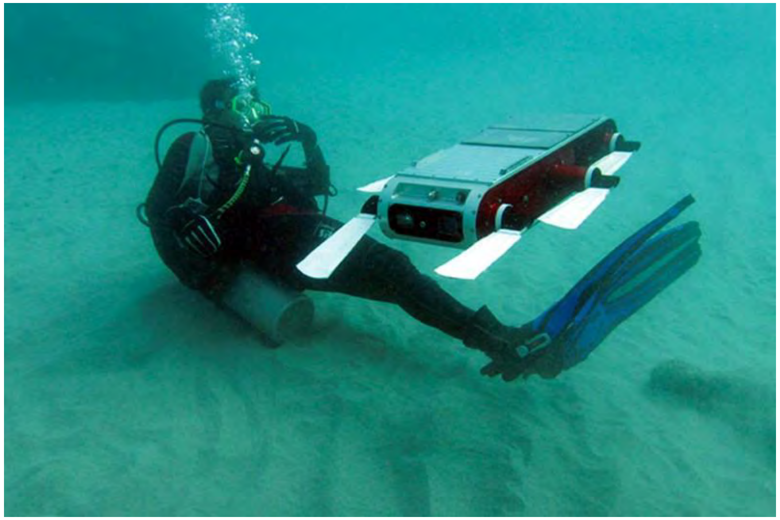
- Manual studies: costly in time and resources, non-reproducible, limited...
- Solution: Methods based on video acquisition

Sébastien Villon, LIRMM marbec Gérard Subsol

David Mouillot, Sébastien Villeger, Thomas Claverie

Localization / identification

(D. Mouillot, S. Villeger, T. Claverie, S. Villon, G. Subsol, M. Chaumont)



Video analysis allows:
- to increase the volume of recovered data,
- to verify and compare the results,
- to overcome human physical limitations.

Localization / identification

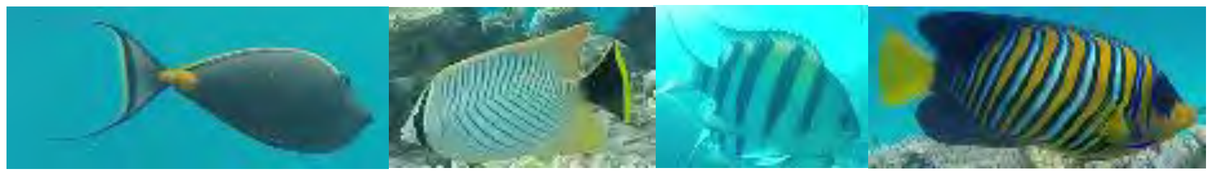
(D. Mouillot, S. Villeger, T. Claverie, S. Villon, G. Subsol, M. Chaumont)

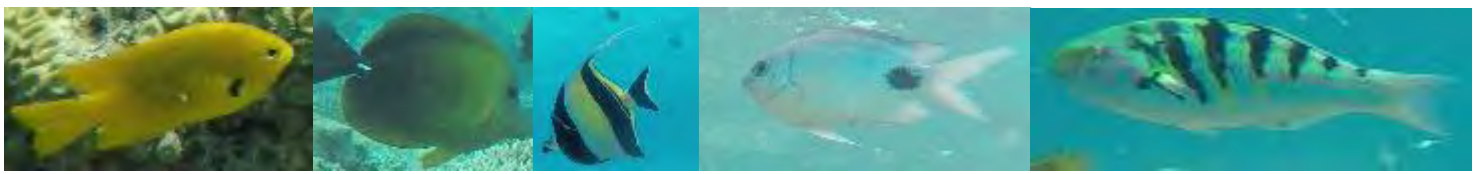
How much is it reliable?

(D. Mouillot, S. Villeger, T. Claverie, S. Villon, G. Subsol, M. Chaumont)

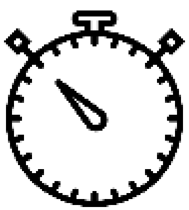
Let us compare !

9 species





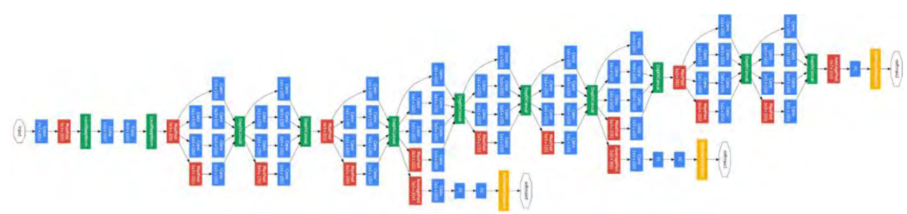
The Human VERSUS The Machine



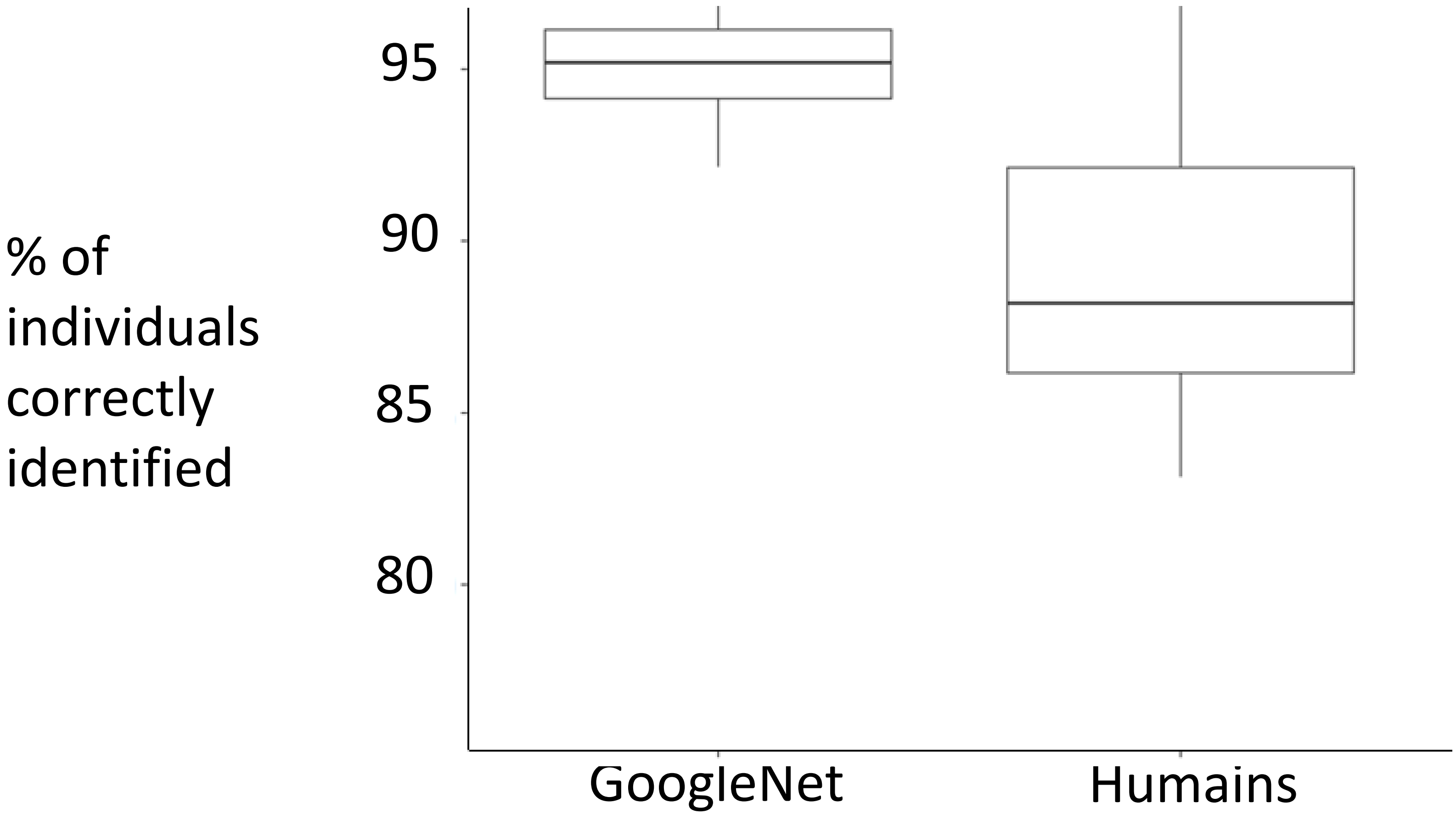
20 minutes / Human

14 Humans

GoogleNet
(trained on 20 species)



Result



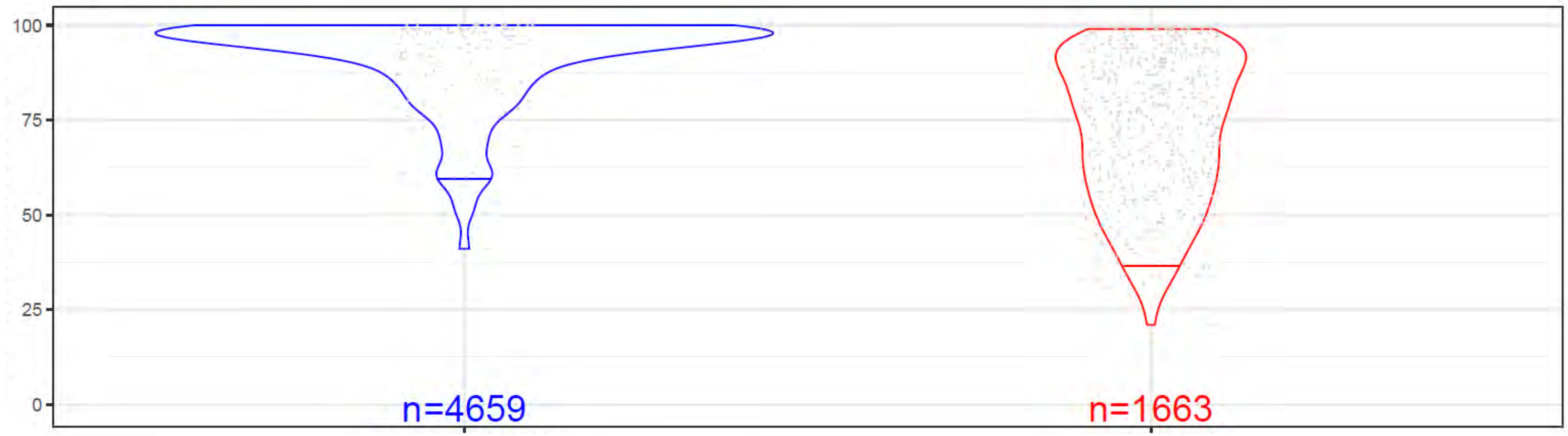
The machine is **6% more accurate and 100 times faster**

Villon, Sébastien, et col. "**A Deep learning method for accurate and fast identification of coral reef fishes in underwater images**." *Ecological informatics* 48 (2018).

Can we really trust the results?

Output scores on a test set

Scores output by the network



Case where it is a correct classifications

Case where it is an **<u>in</u>**<u>correct</u> classification

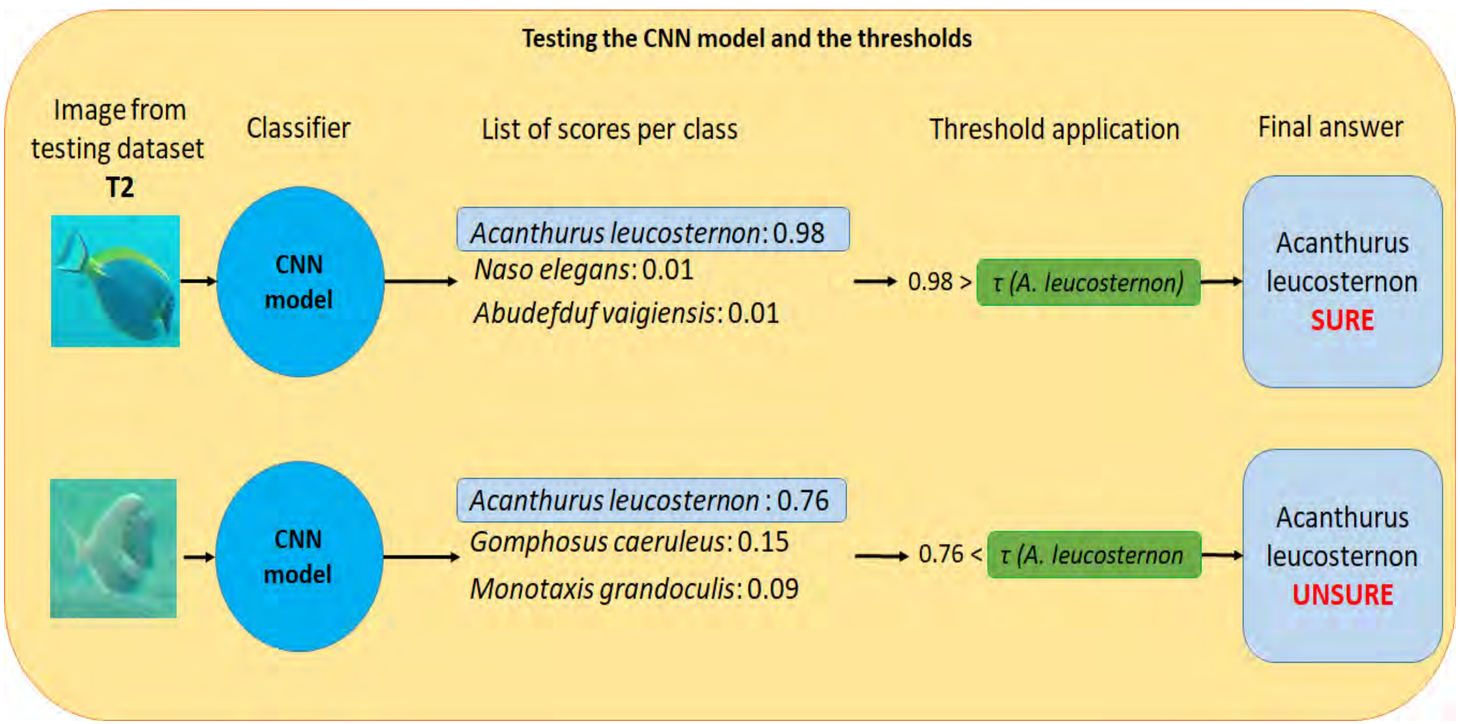
A possible solution

The post-processing:

The network can:

- predict a species,

-or can **refuse to predict ("unsure class")**

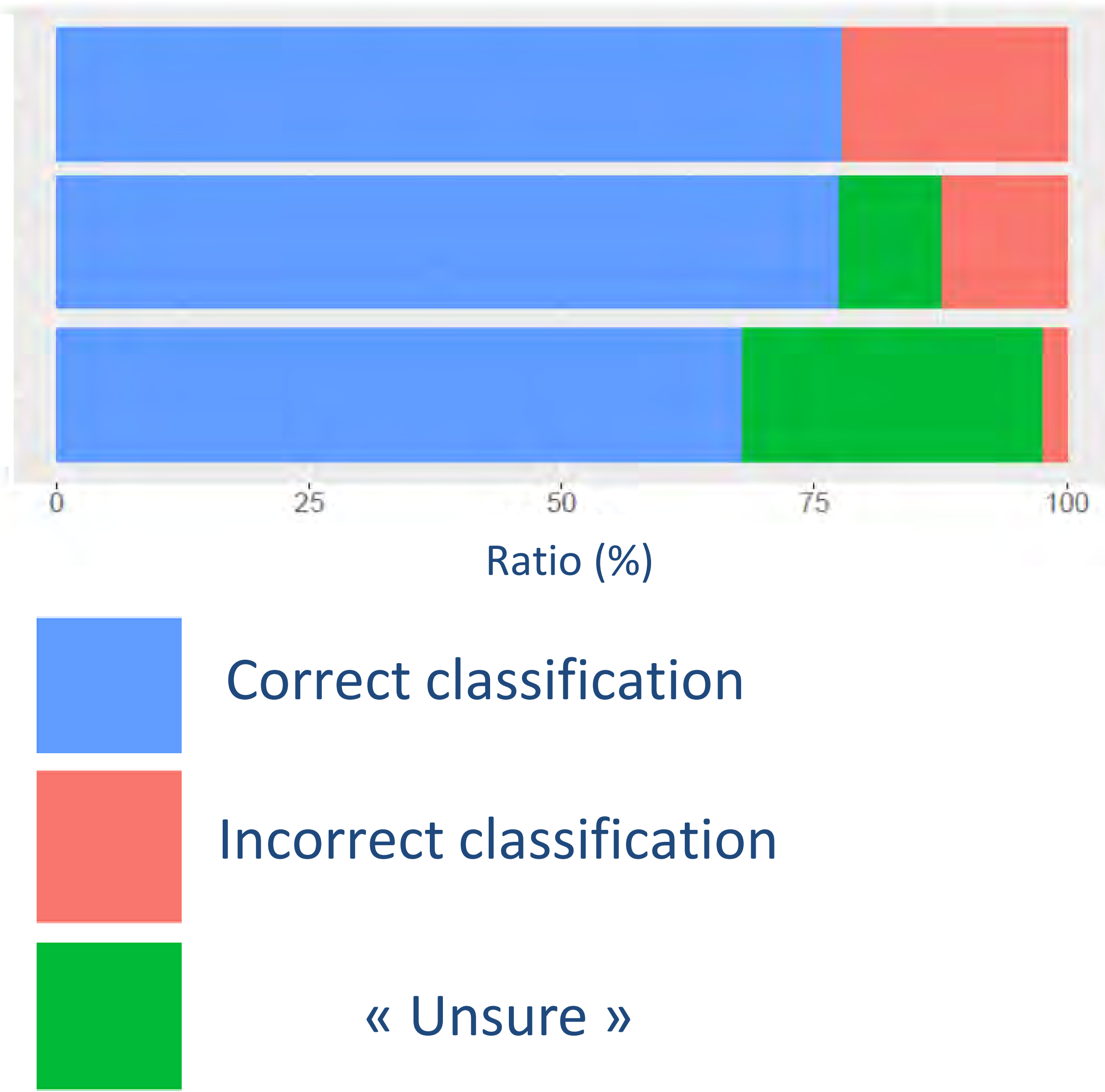


Results

No post-processing

Objective n°1:
Keep correct classification maximum

Objective n°2:
Minimize the incorrect classification

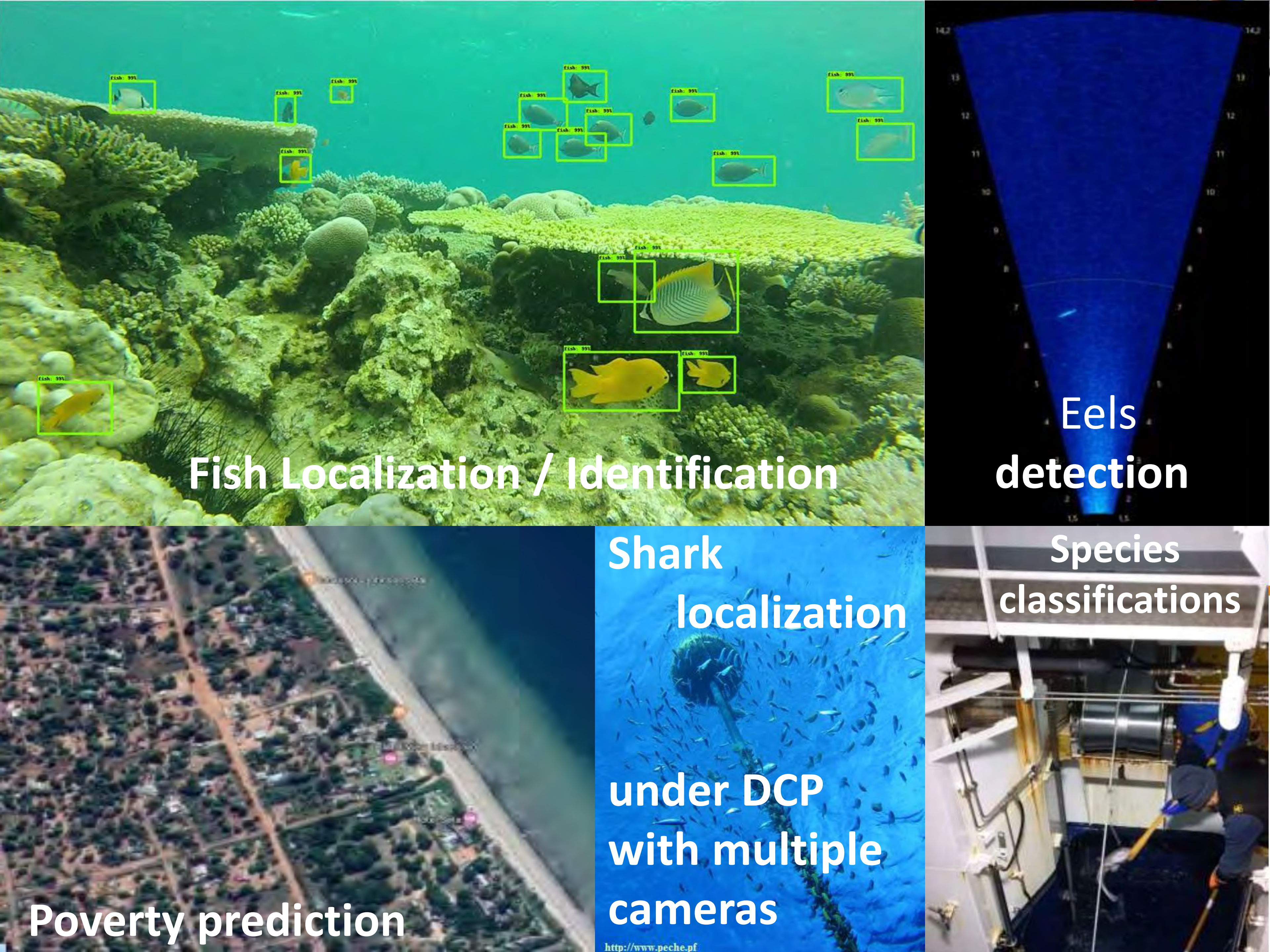


New project (in continuity)

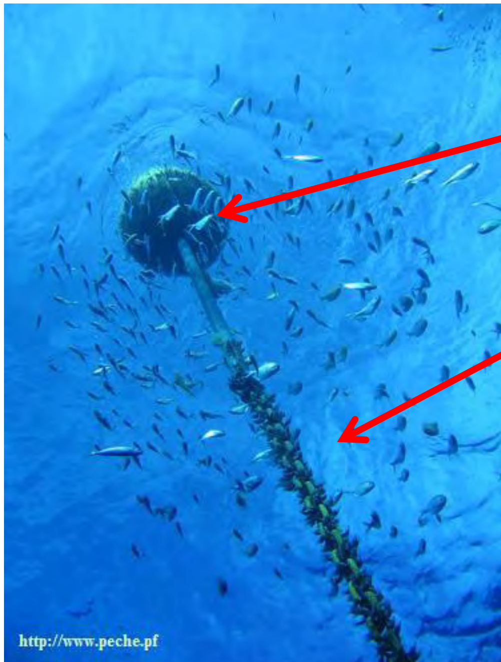
OBJECTIVE:
- Better detection,
- Size estimation,
- 3D motion estimation,
→ Behavior estimation

Stereo-System……………………….fixed on a…………robot

" Detection and counting of fish with a stereo-vision system",
research engineer: Sep. 2020 – August 2021 @LIRMM, Montpellier, ICAR.
Supervisors: Marc Chaumont, Gérard Subsol, …


Fish Localization / Identification
Eels detection
Poverty prediction
Shark localization under DCP with multiple cameras
http://www.peche.pf
Species classifications

Shark localization // multi-view



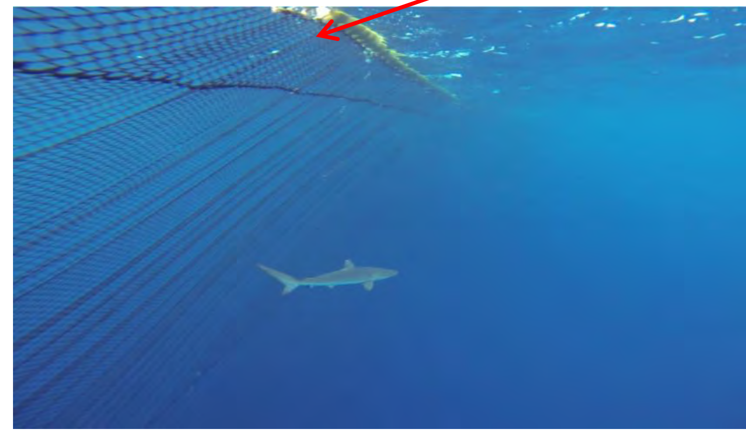
Fish Aggregating Device (FAP)

Chain

Container with the camera<u>s</u>

Prototype :
(this is just an « artistic » vision…)

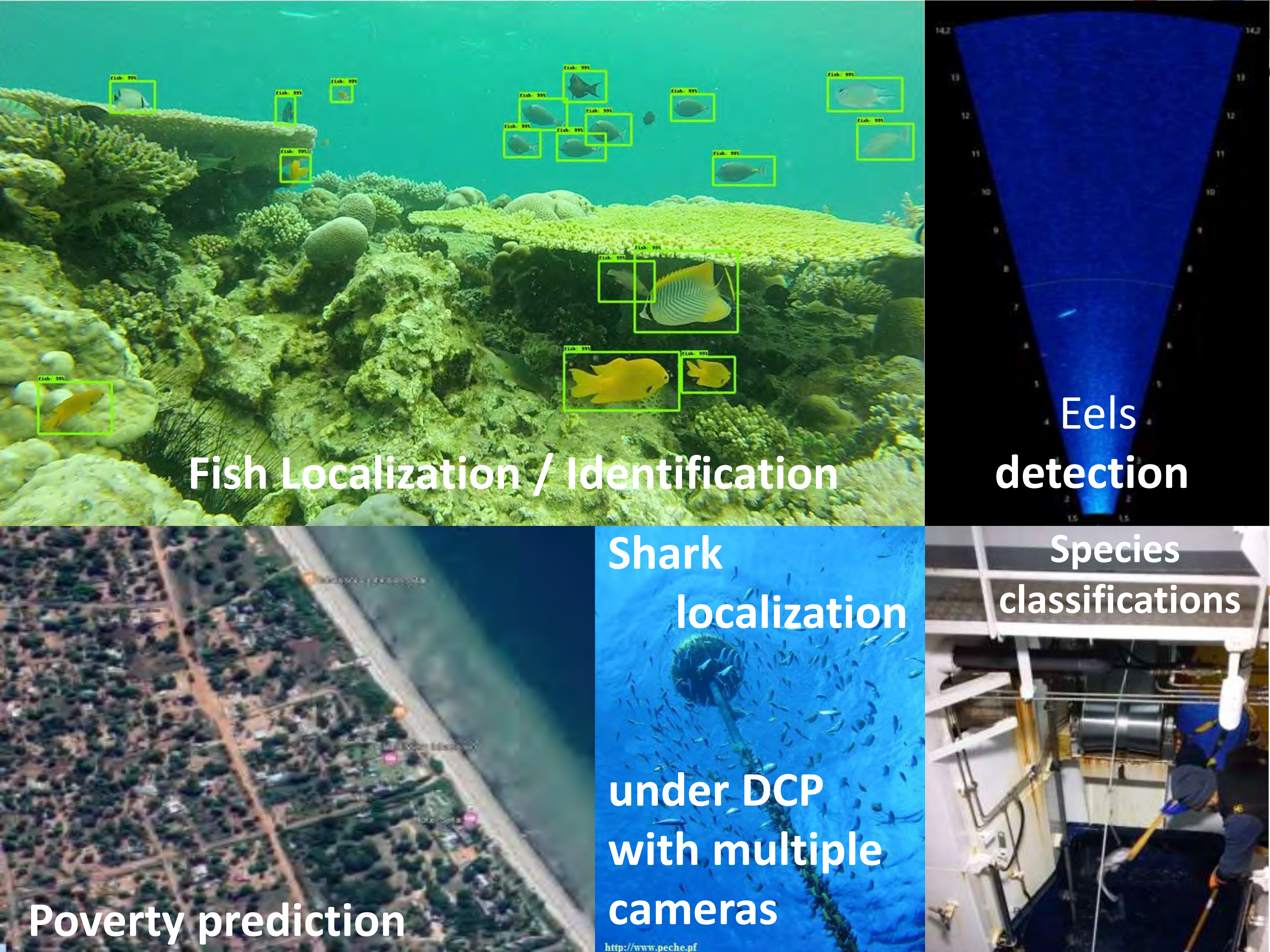
fishing net



Shark

Tunas

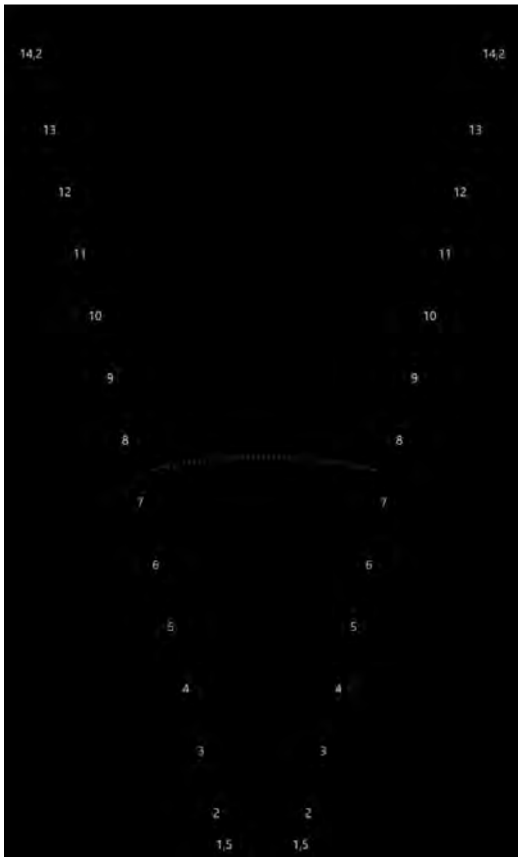
" Detection and counting of sharks from a sequence of multi-view underwater images by Deep-Learning ",
Post-Doc: Mehdi Yedroudj : <u>Dec. 2019</u> – <u>August 2020</u> @LIRMM, Montpellier, ICAR.
Supervisors: Marc Chaumont, Gérard Subsol, Vincent Creuze, Laurent Dagorn, Manuela Capello

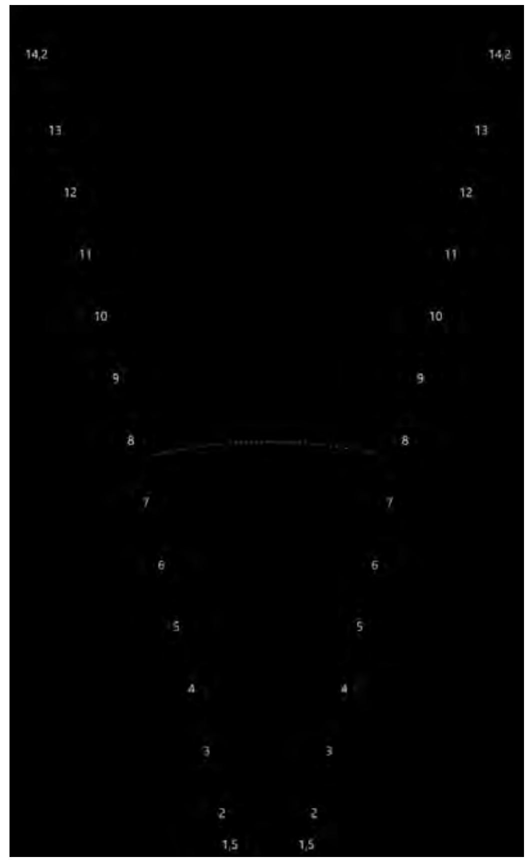


Fish Localization / Identification
Eels
detection
Shark
localization
under DCP
with multiple
cameras
http://www.peche.pf
Species
classifications
Poverty prediction

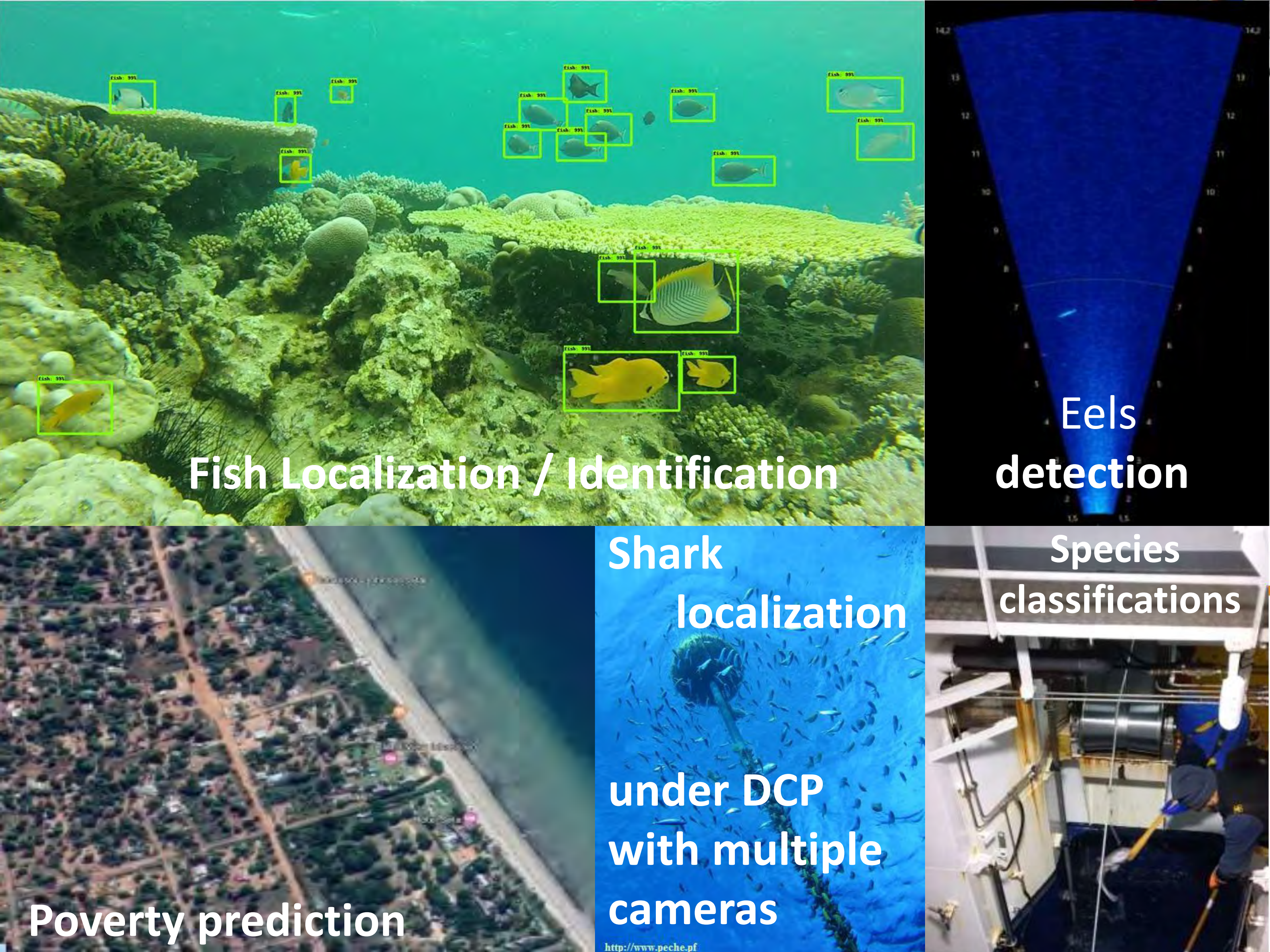
Eels counting

European management plan for the restoration of the eel stock





" [Identification and counting of eels from multibeam sonar videos by Deep-Learning]",
Master 2 internship **2020** @LIRMM, Montpellier, ICAR.
Supervisors: Gérard Subsol, Vincent Creuze, Mehdi Yedroudj, Marc Chaumont, Jason Peyre, Raphaël Lagarde

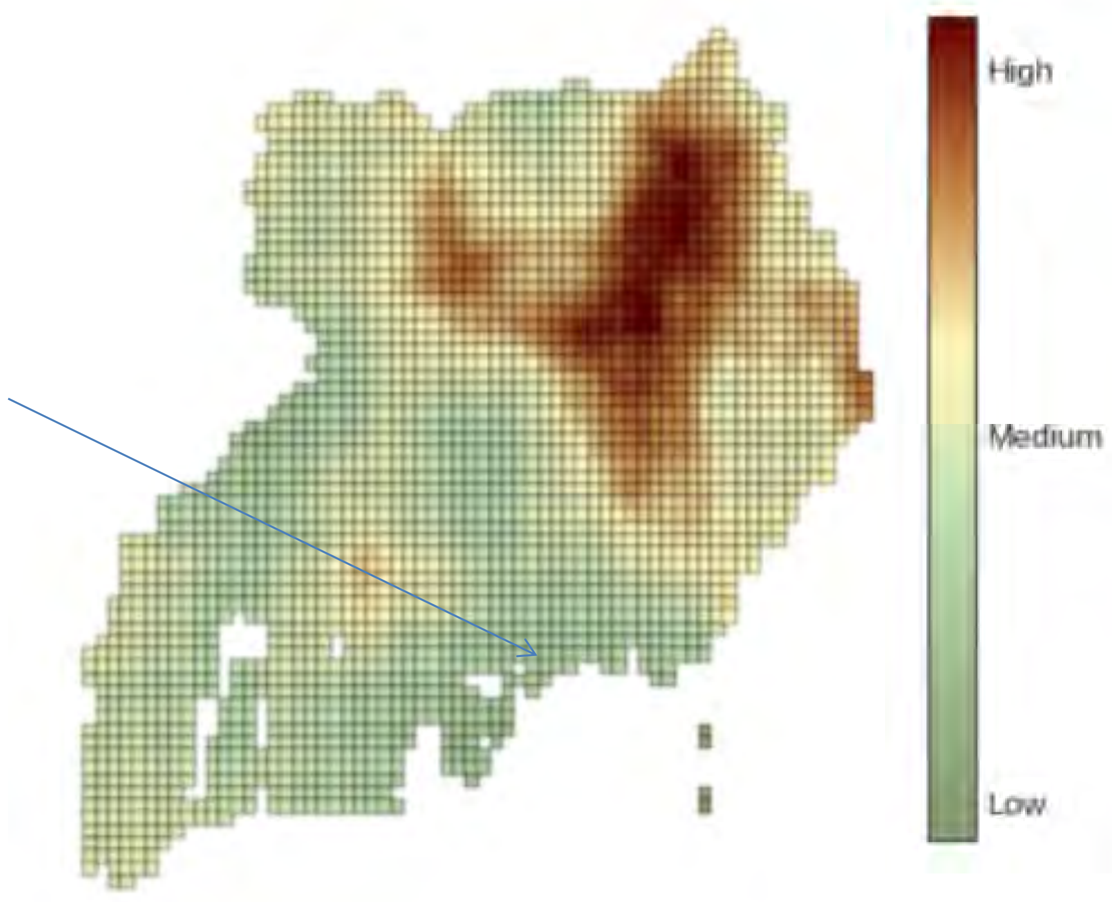

Fish Localization / Identification
Eels
detection
Shark
localization
under DCP
with multiple
cameras
http://www.peche.pf
Species
classifications
Poverty prediction

Prediction of poverty with a CNN ?

From a 400x400 pixel images (1 km x 1km) the CNN should predict a poverty value (scalar) (0 = Low ; 100 = high)

Ask to Google Static Maps API, for the image 400 × 400 pixels at zoom level 16

Predicted poverty in Uganda
(poverty ≈ annual consumption level of households)



Predicted poverty probabilities at a fine-grained 10km × 10km block level.

Image from « Transfer learning from deep features for remote sensing and poverty mapping » M. Xie et al. AAAI’2016

Is it done?
Is there still room for <u>computer-science</u> research?

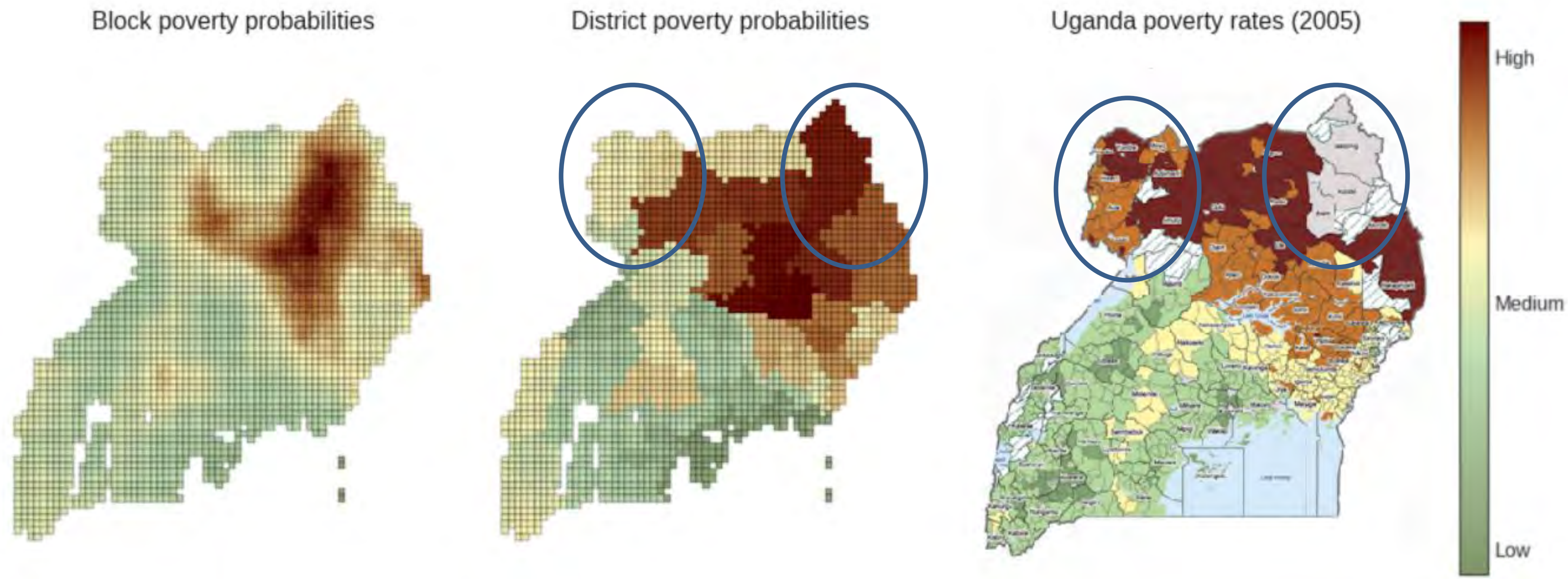


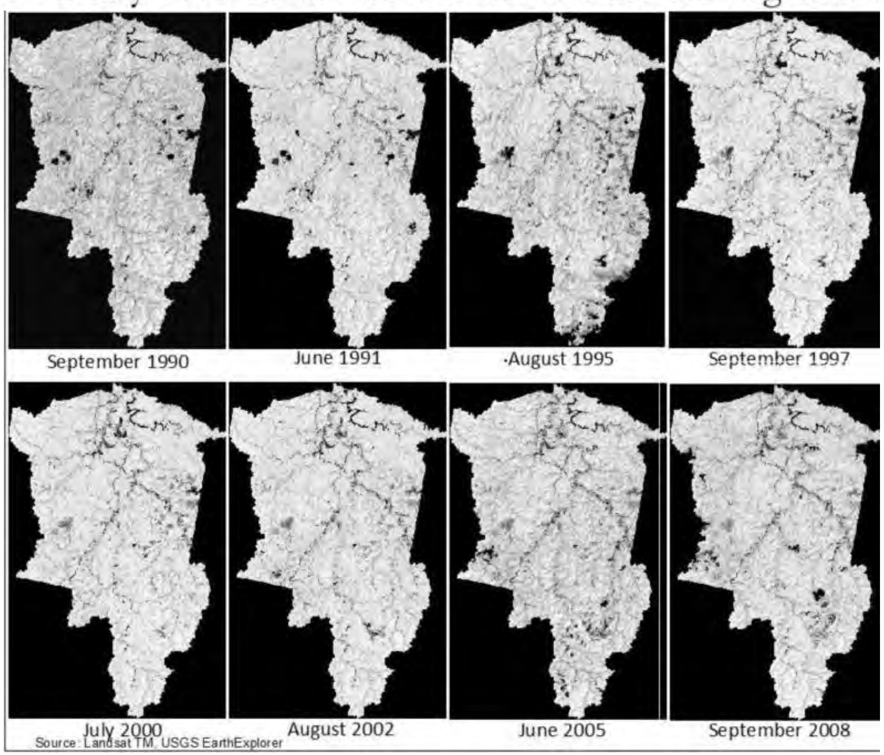
Figure 3: **Left:** Predicted poverty probabilities at a fine-grained 10km × 10km block level. **Middle:** Predicted poverty probabilities aggregated at the district-level. **Right:** 2005 survey results for comparison (World Resources Institute 2009).

Only correlated to 70% to the ground truth

Image from « Transfer learning from deep features for remote sensing and poverty mapping » M. Xie et al. AAAI'2016

Objective: work with images sequences

Floyd County 1990-2008 Normalized Difference Vegetation Index



https://ericsvenson.com/monitoring-mining-impact/

Deep learning → Poverty (0% … 100%)

… variable resolution,
temporal irregularities,
small database,
etc…

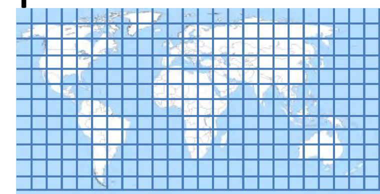
Post-doc (funded by Belmont/CESAB) 2020

Thesis (ANR) in September 2020

" **Poverty prediction by Deep Learning from a satellite image sequence** ",
Master 2 internship <u>2020</u> @LIRMM, Montpellier, ICAR.
Supervisors: Marc Chaumont, Gérard Subsol, Laure Berti-Équille, Dino Ienco

Other projects …

- Prediction of the number of pelagic species for a GPS position
  Post-doc : Laura Mannocci (Marbec)
  Start in January 2020.



- Prediction of genomic hybridization of European brown trout by image analysis (Marbec)
  Master 1 internship <u>2020</u>

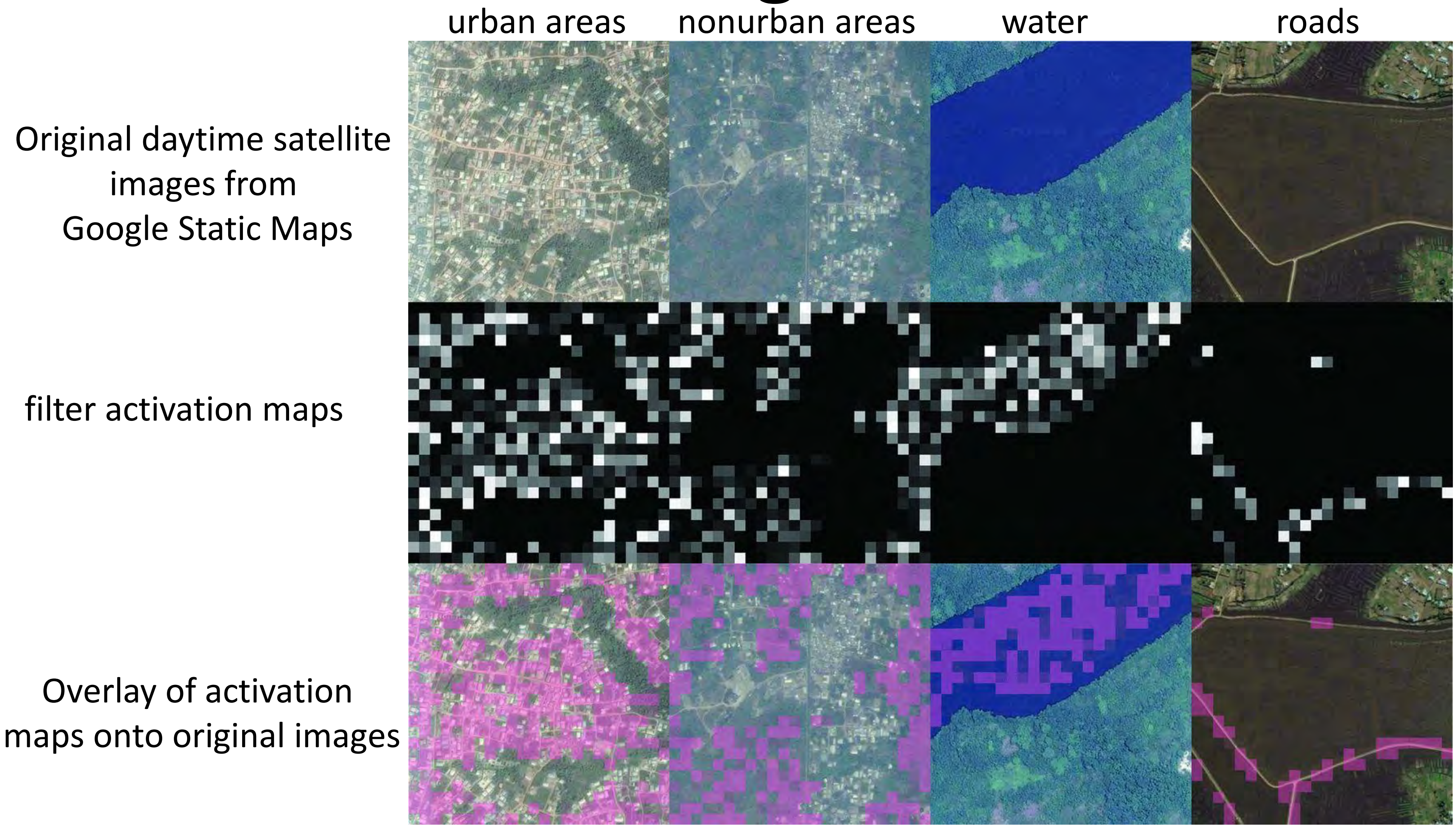
- Evaluation of wildlife crossings (bridges and tunnels) along highways (with VINCI Autoroutes, France and Claude MIAUD@ CEFE CNRS)
  Master internship 2021 ?

<u>SUBJECTS discussed in the past … :</u>

- **Detection of plastics** on the surface and in water (Marbec, Sète, France),
- Identification of **fish, gorgonians and algae** (Banyuls Observatory, France),
- Detection of **fish malformations** ("Poissons du soleil", Balaruc, France),
- Study of **fish larvae** (ECOCEAN society),
- **DNA analysis/comparison** of marine species (SPYGEN society, France),
- Analysis of **microscopic images of coral** reproduction (CORAIL Laboratory - CRIOBE Moorea),
- TAAF…

TO BE CONTINUED ...

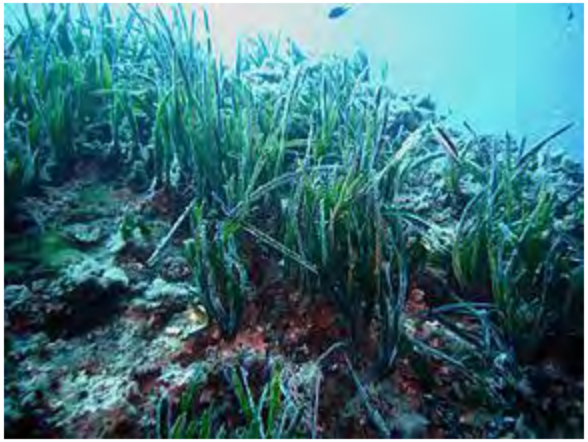
Parts of the image that “react”



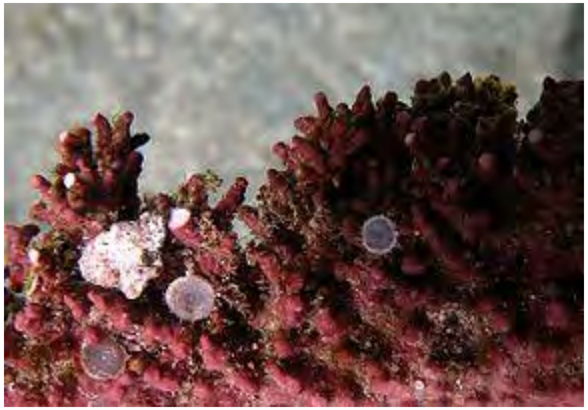
[Combining satellite imagery and machine learning to predict poverty]() N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, S. Ermon, Science, 353(6301), 790-794, 2016

Comptage du nombre de bateaux

- La baie de Paulilles (aire protégée)



Posidonies

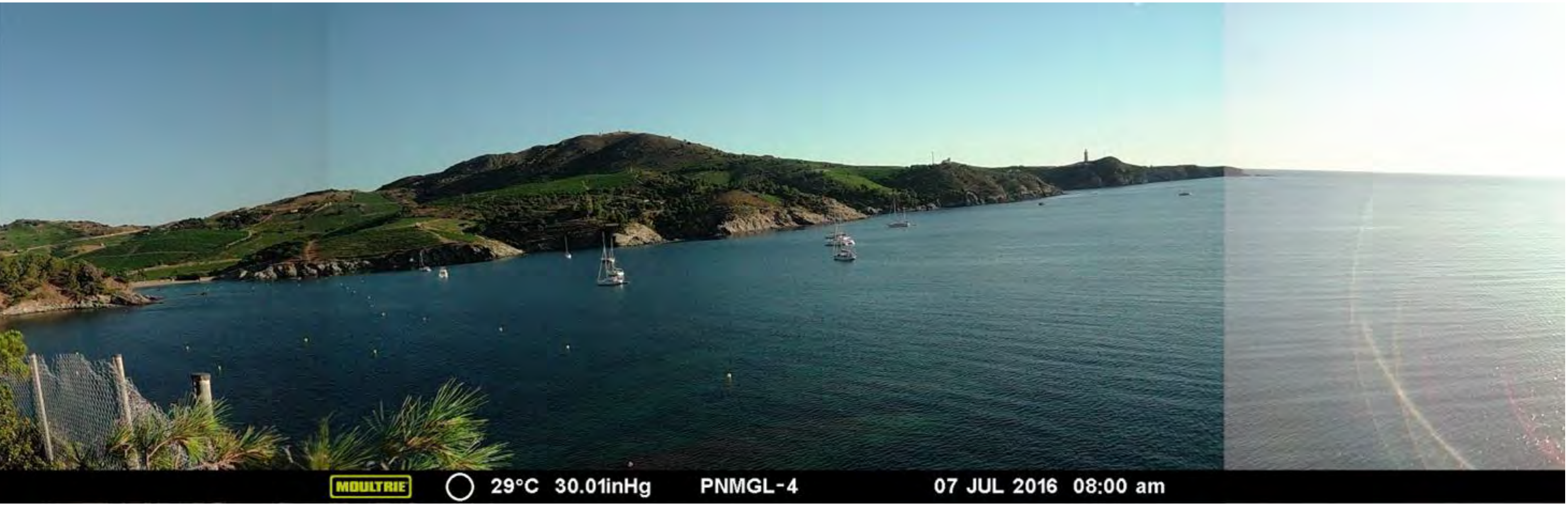


Coralligène = éco-syst. avec algues calcaire

- Question du parc marin :
  Corrélation nb bateau / dégradation fond marin ?

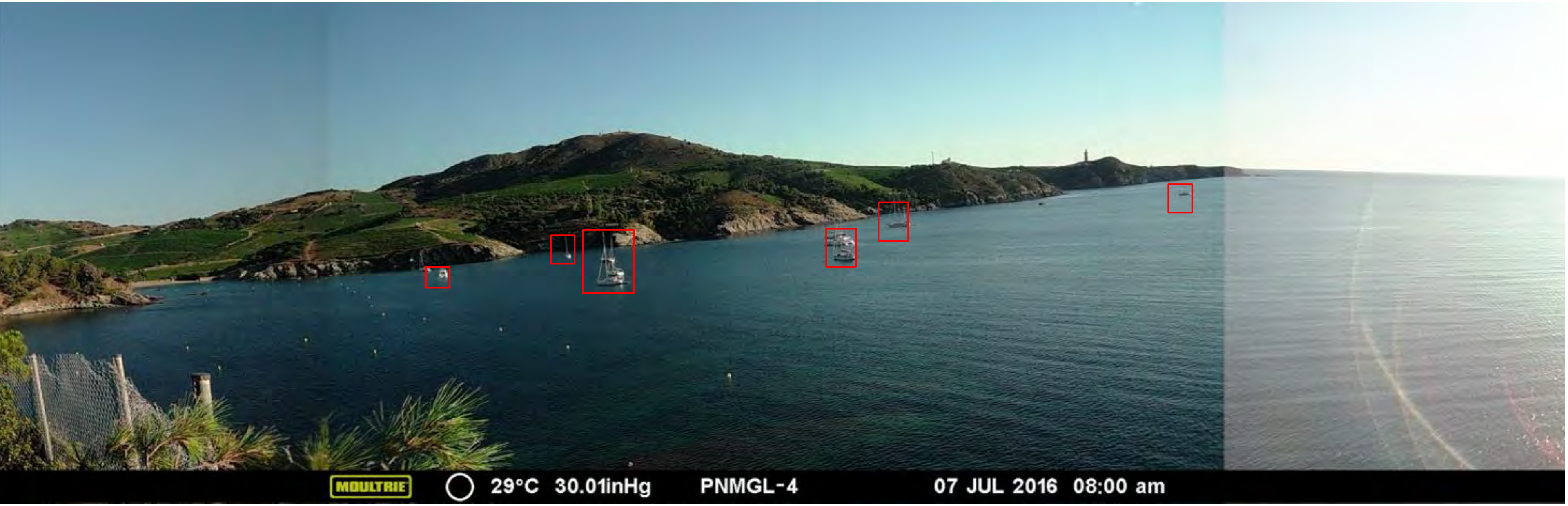
WILLIAM Puech, VINCENT Itier, GRAPTIN Régis, Younes Zegaoui

Dispositif de surveillance



- Image de 10 656 x 1 998 pixels
- Un image toute les 30 minutes

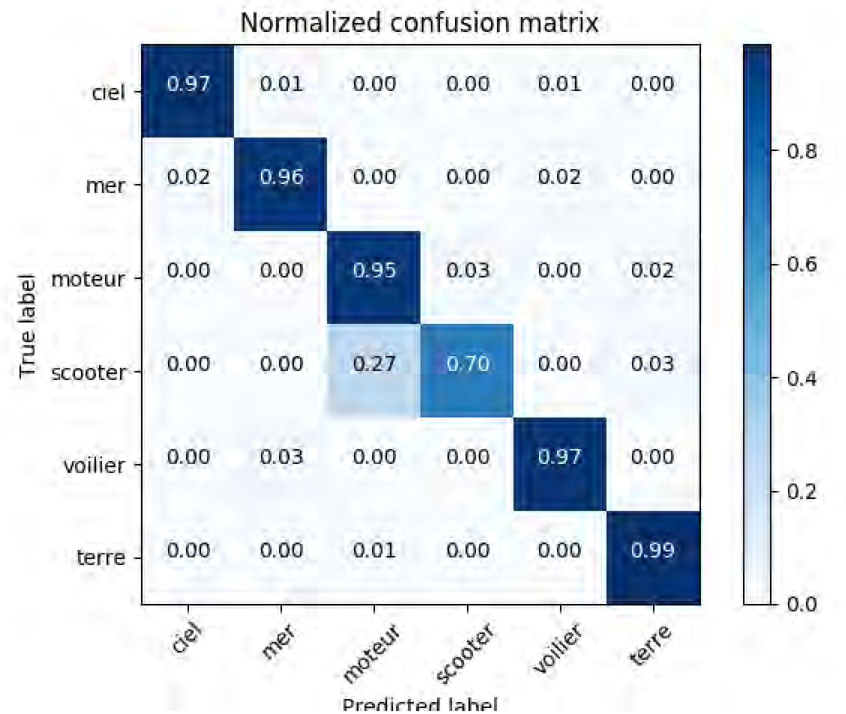
Ce que l'on voudrait obtenir



Etape préliminaire (base + classif)

	Entraînement	Validation
Ciel	**2094**	**518**
Mer	**2086**	**526**
Moteur	**1952**	**252**
“Scooter”	**576**	**74**
Terre	**1605**	**552**
Voilier	**2183**	**440**

Base de donnée d’apprentissage



Matrice de confusion sur base de validation (2362 images).

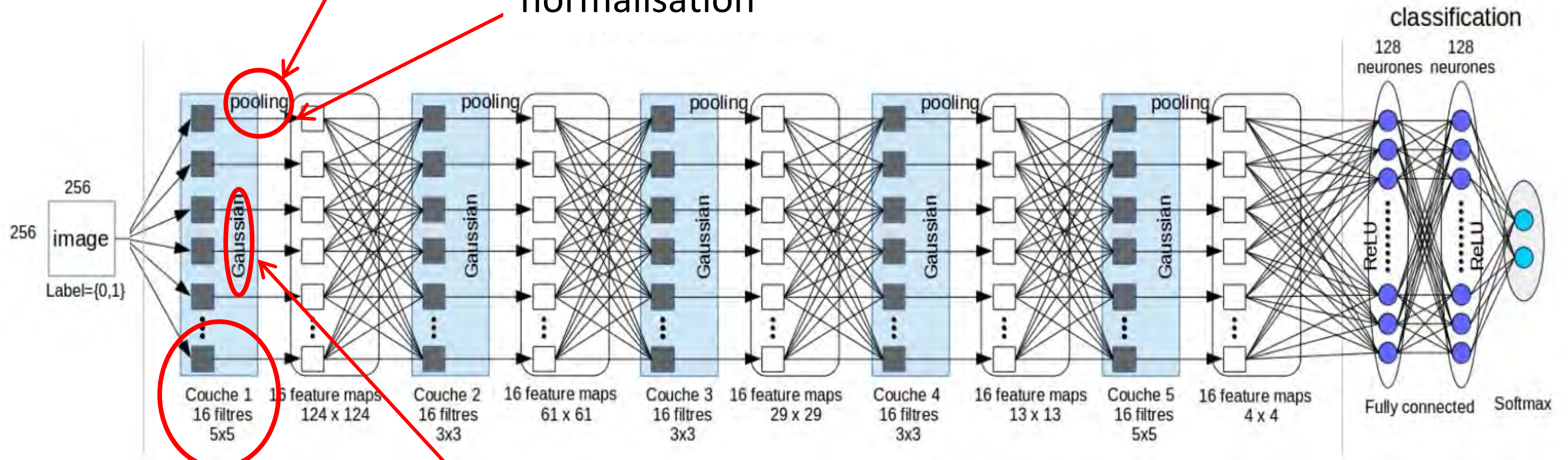
Inception v3 pré-entrainé sur 2 millions d’images sur ImageNet (tranfer learning)


That’s all folks!

Le réseau CNN
Convolutional Neural Network

moyenne ou max [+ sous-échantillonage]

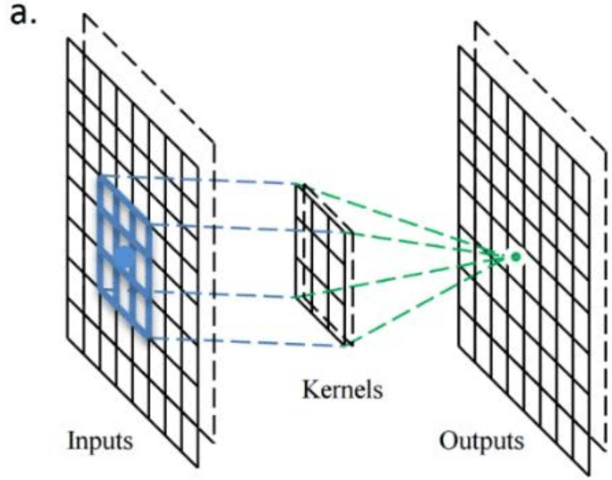
normalisation



convolutions

fonction non linéaire
(= fonction d'activation)

La convolution



Exemple : convolution d'une zone 3x3 de l'image avec un « kernel »

200	210	15
255	180	7
100	63	0

contenu local de l'image

1	1	1
1	1	1
1	1	1

« kernel » (noyau)

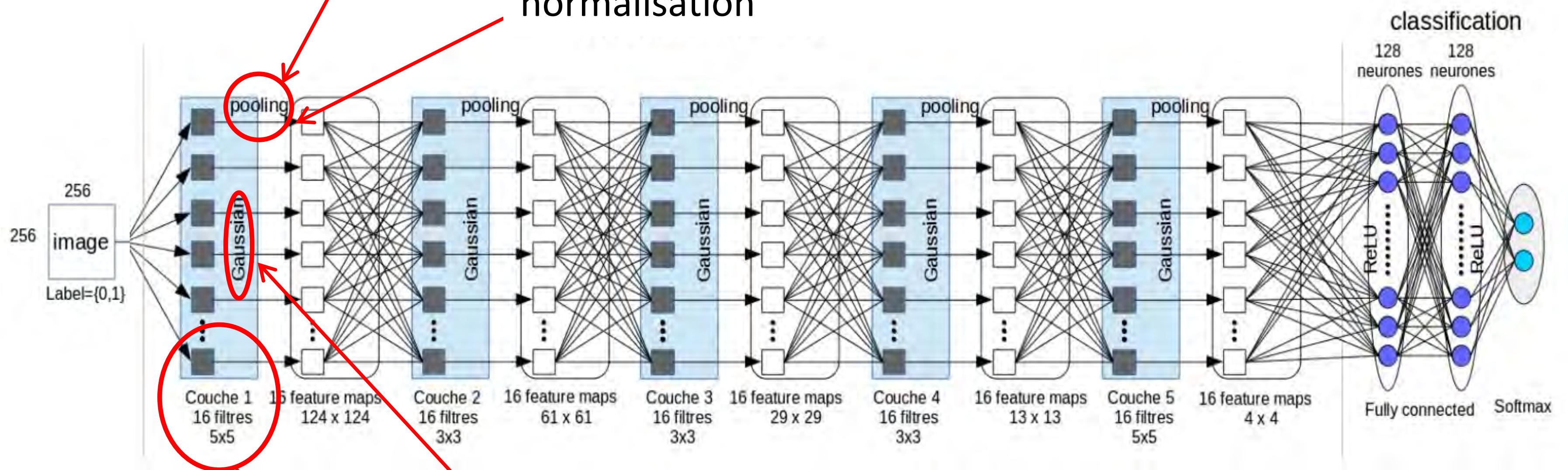
?	?	?
?	114,4	?
?	?	?

résultat (output)

Le réseau CNN
Convolutional Neural Network

moyenne ou max [+ sous-échantillonage]

normalisation

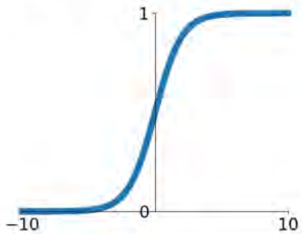


convolutions

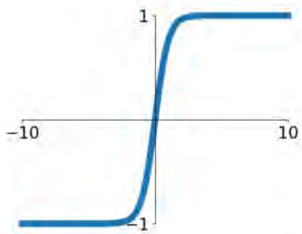
fonction non linéaire
(= fonction d'activation)

Activation Functions

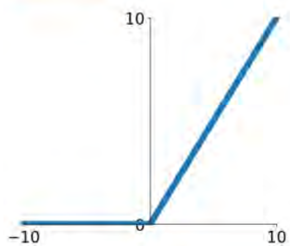
Sigmoid
$\sigma(x) = \frac{1}{1+e^{-x}}$



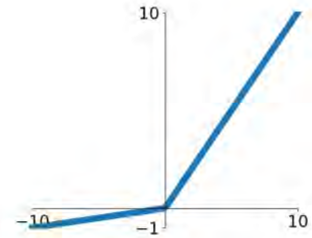
tanh
$\tanh(x)$



ReLU
$\max(0, x)$

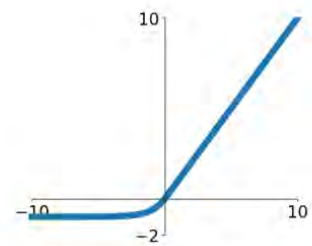


Leaky ReLU
$\max(0.1x, x)$



Maxout
$\max(w_1^T x + b_1, w_2^T x + b_2)$

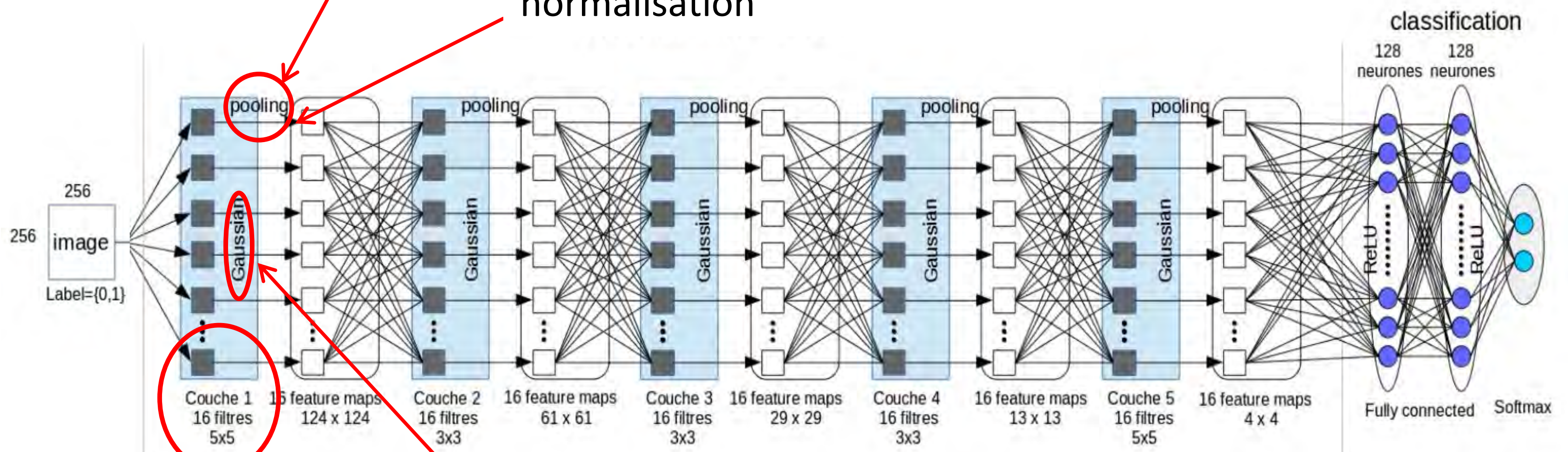
ELU
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



Le réseau CNN
Convolutional Neural Network

moyenne ou max [+ sous-échantillonage]

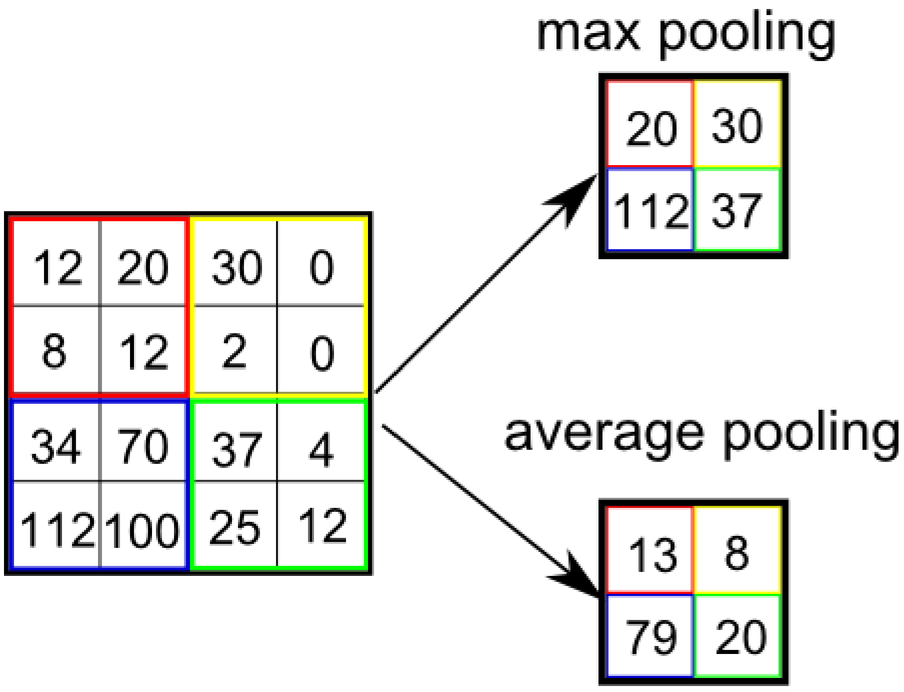
normalisation



convolutions

fonction non linéaire
(= fonction d'activation)

Le pooling

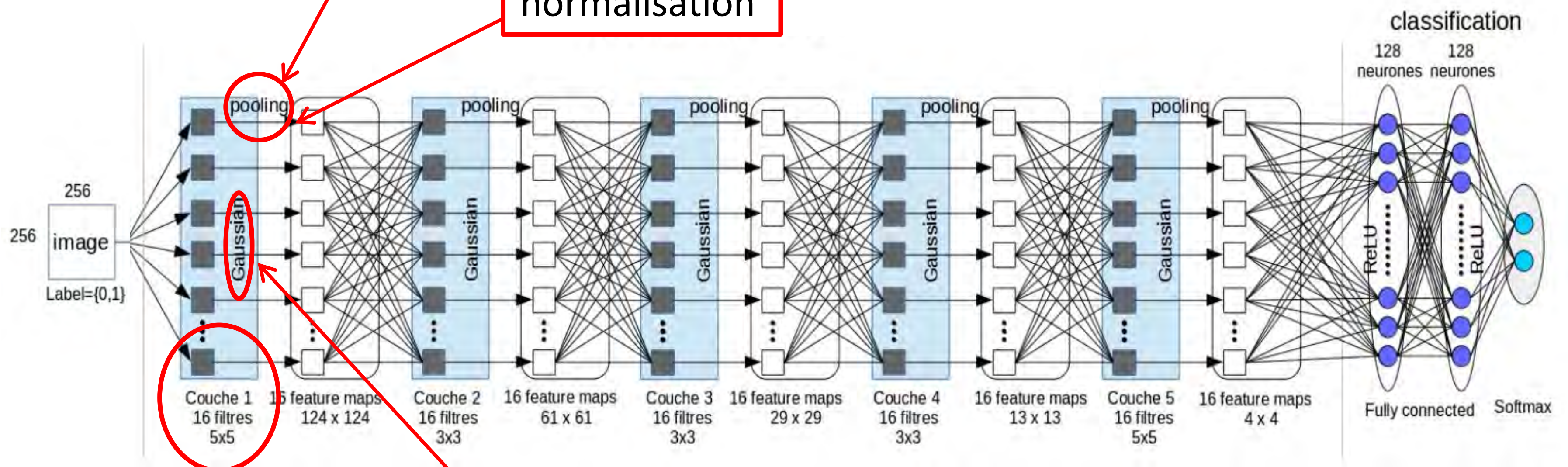


12
20
30
0
8
12
2
0
34
70
37
4
112
100
25
12
max pooling
20
30
112
37
average pooling
13
8
79
20

Le réseau CNN
Convolutional Neural Network

moyenne ou max [+ sous-échantillonage]

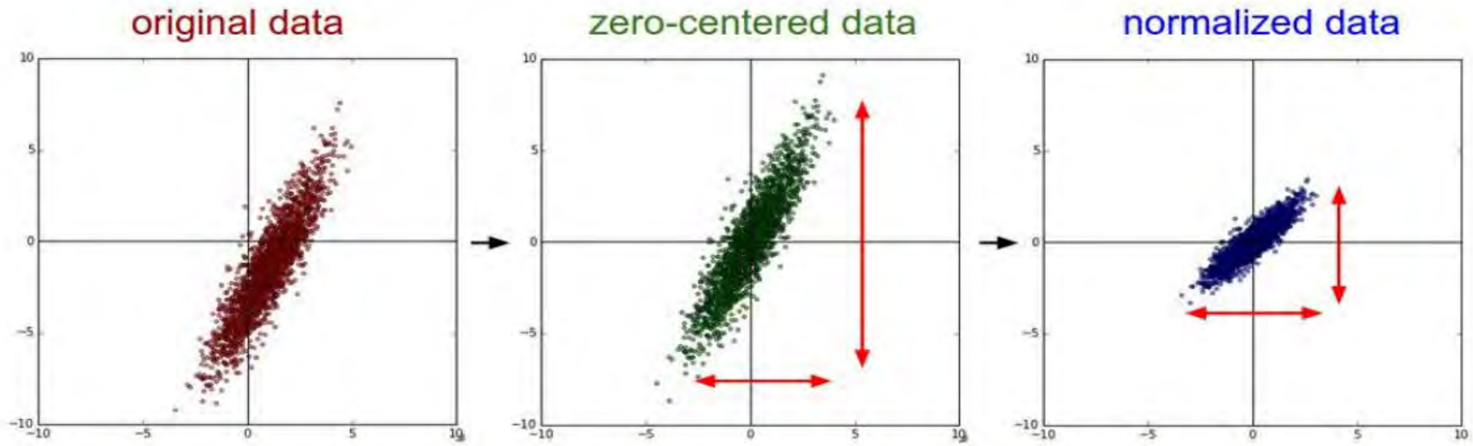
normalisation



convolutions

fonction non linéaire
(= fonction d’activation)

Exemple : Batch Normalisation



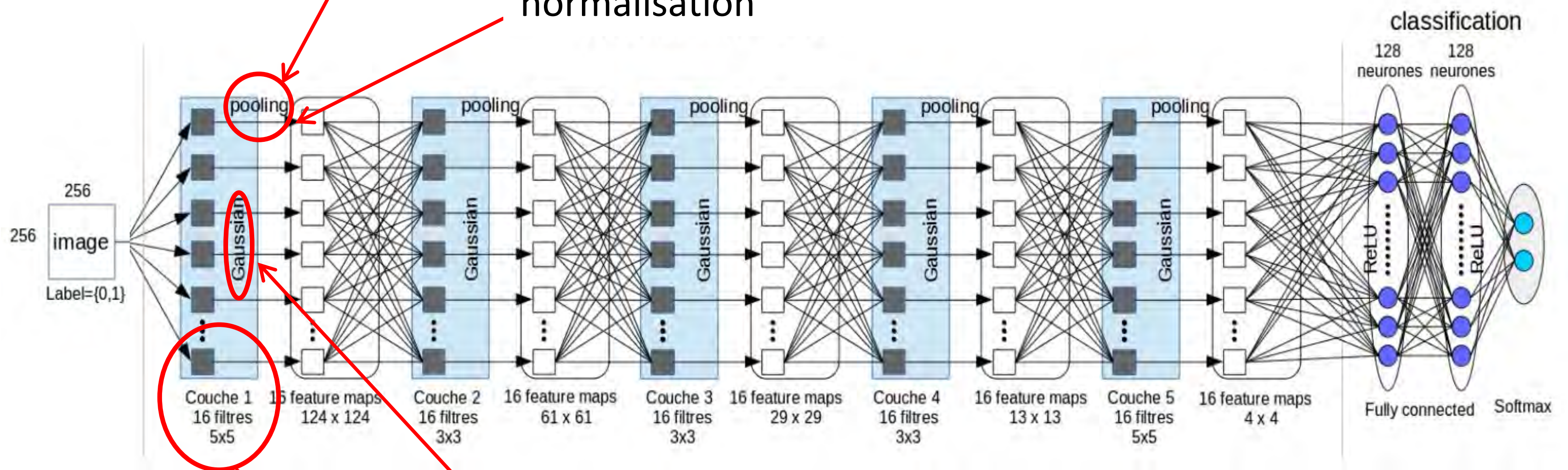
Batch Normalization

$$BN(X, \gamma, \beta) = \beta + \gamma \frac{X - E[X]}{\sqrt{Var[X] + \epsilon}}, \text{ [3][5]}$$

Le réseau CNN
Convolutional Neural Network

moyenne ou max [+ sous-échantillonage]

normalisation



convolutions

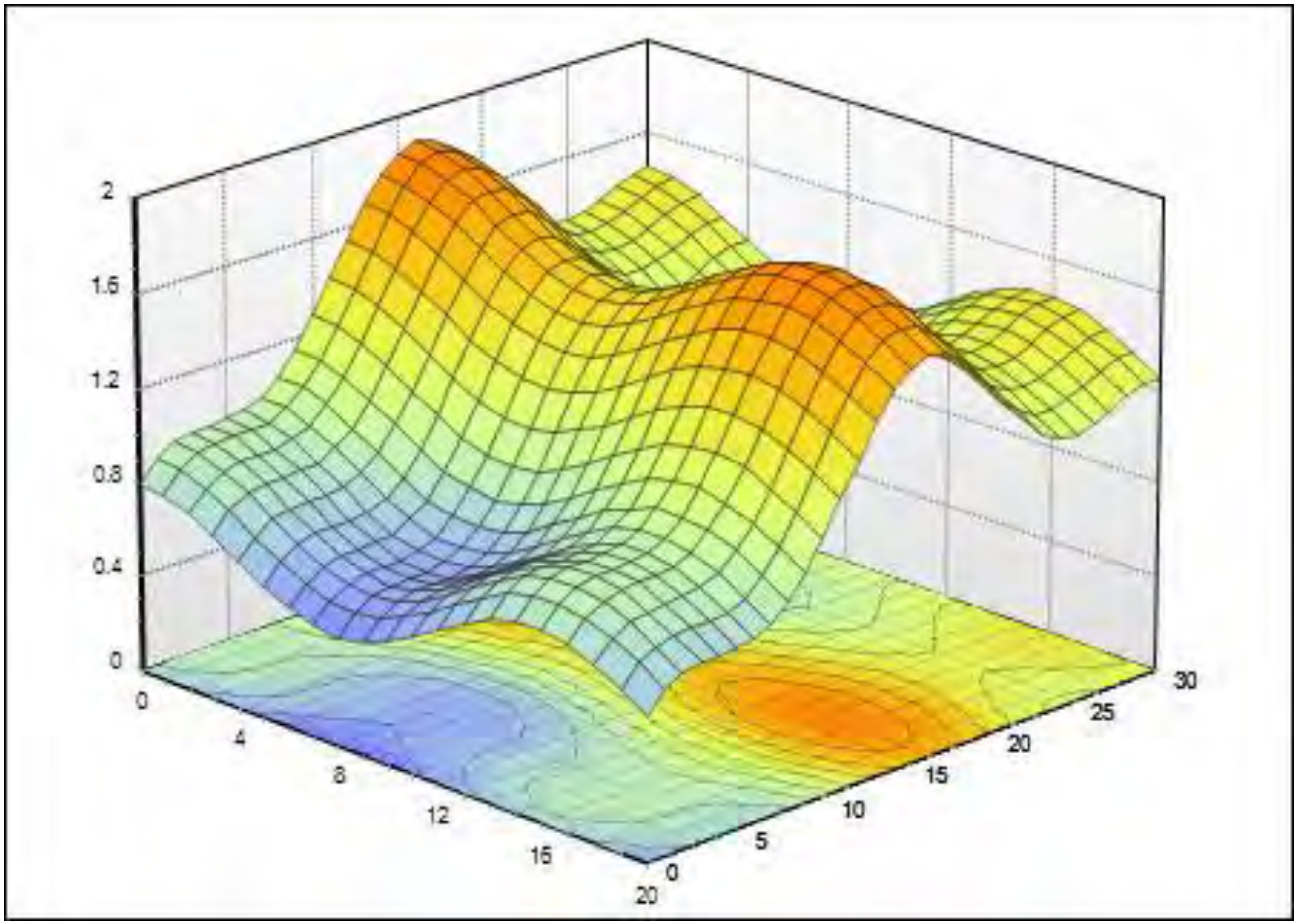
fonction non linéaire
(= fonction d'activation)

Formellement
(et grossièrement..)

$$I_k^{(l)} = norm\left(pool\left(f\left(b_k^{(l)} + \sum_{i=1}^{i=K^{(l-1)}} I_i^{(l-1)} \star F_{k,i}^{(l)}\right)\right)\right)$$

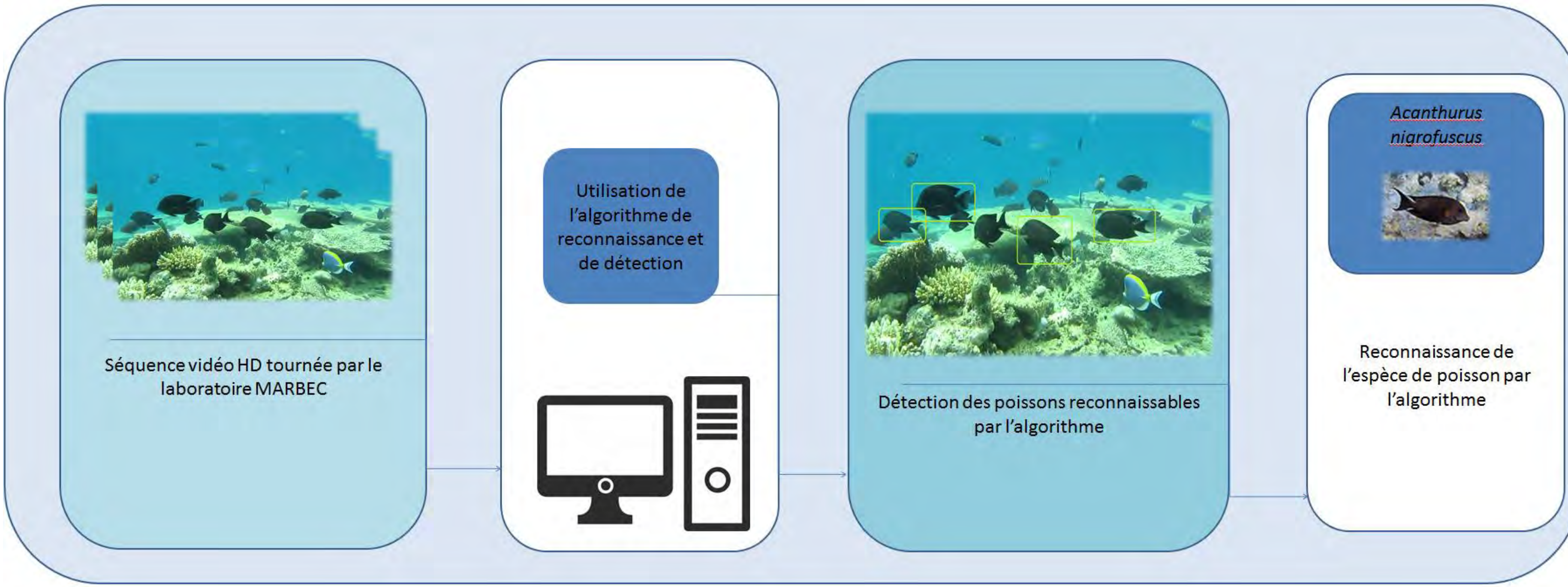
L’apprentissage
(expliqué avec les mains..)

Surface représentant la « distance » entre le score donné par le réseau et la vérité terrain



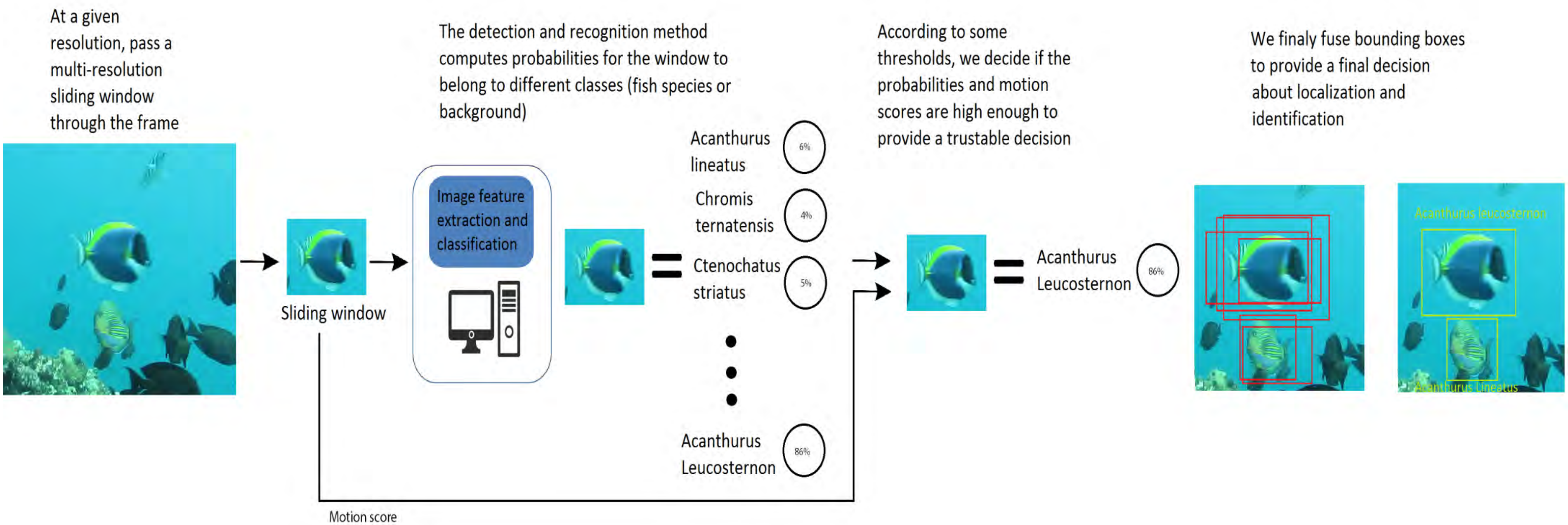
Localisation/Identification automatique de poissons dans des vidéos sous-marines

Le pipeline …



Intérêt : biomasse,
détection nouvelles espèces,
utilisation pour l'analyse comportementale, …

L’architecture globale

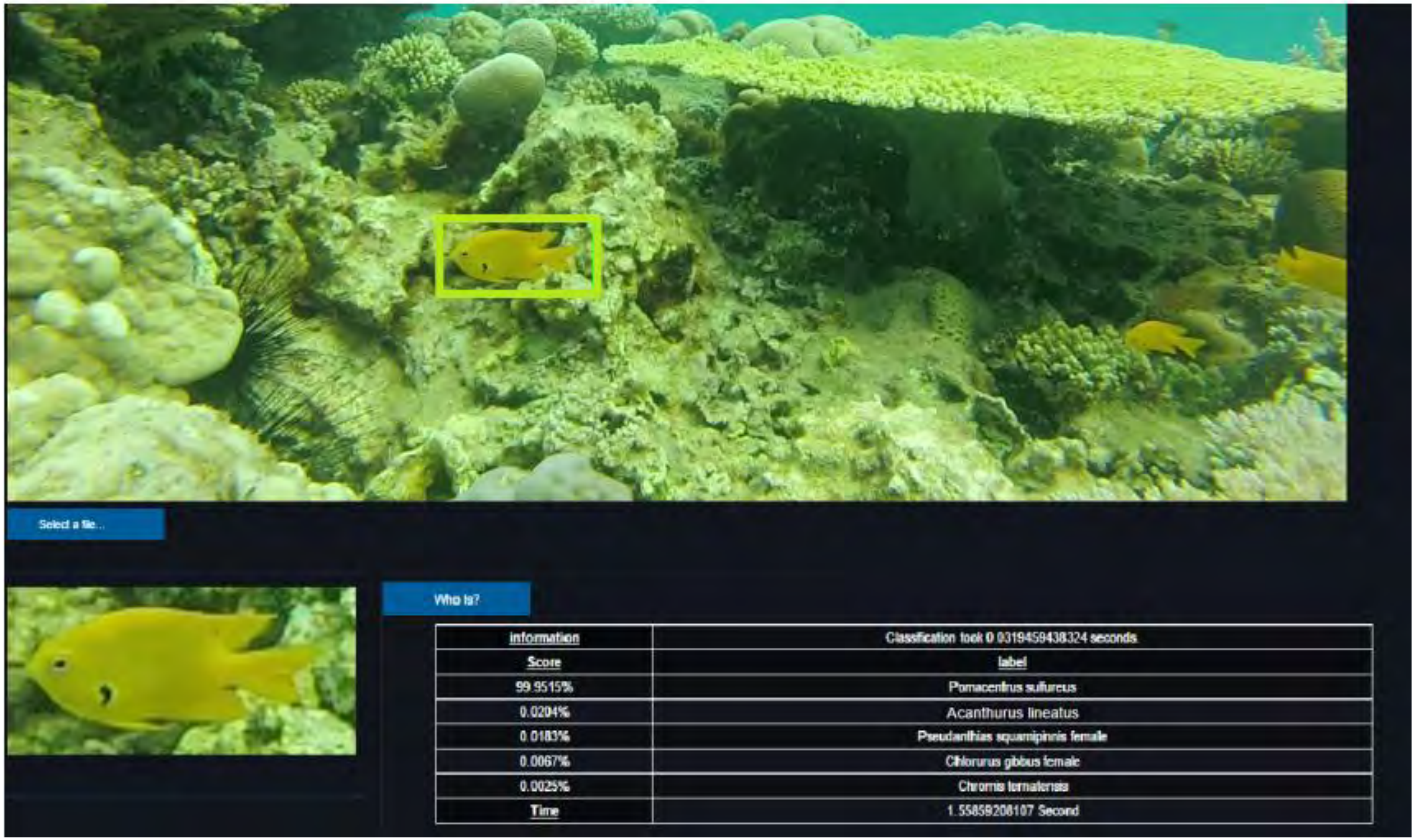


GoogLeNet avec 27 couches et un soft-max.

Sebastien Villon, Marc Chaumont, Gerard Subsol, Sebastien Villeger, Thomas Claverie, David Mouillot, "Coral reef fish detection and recognition in underwater videos by supervised machine learning : Comparison between Deep Learning and hog+svm methods", ACIVS'2016, Advanced Concepts for Intelligent Vision Systems, Lecce, Italy, October 24-27, 2016, 12 pages, published by Springer in the Lecture Notes in Computer Science series.

Une application d’identification

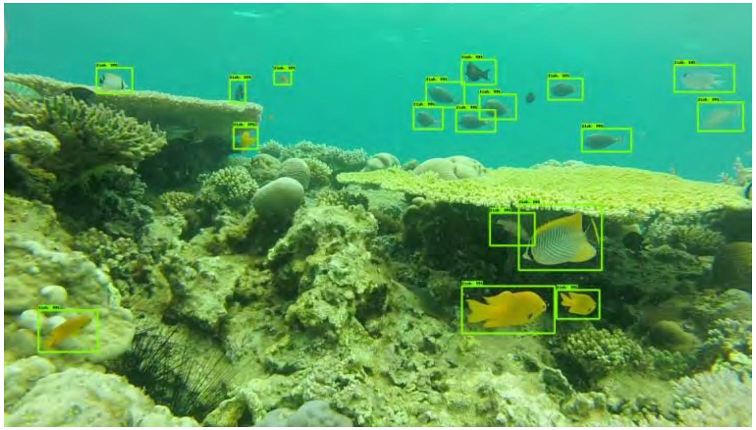
Un utilisateur dessine un rectangle autour du poisson
L’application renvoie un résultat sous forme d’un « top 5 ».



information	Classification took 0.0319459438324 seconds
Score	label
99.9515%	Pomacentrus sulfureus
0.0204%	Acanthurus lineatus
0.0183%	Pseudanthias squamipinnis female
0.0067%	Chlorurus gibbus female
0.0025%	Chromis ternatensis
Time	1.55859208107 Second

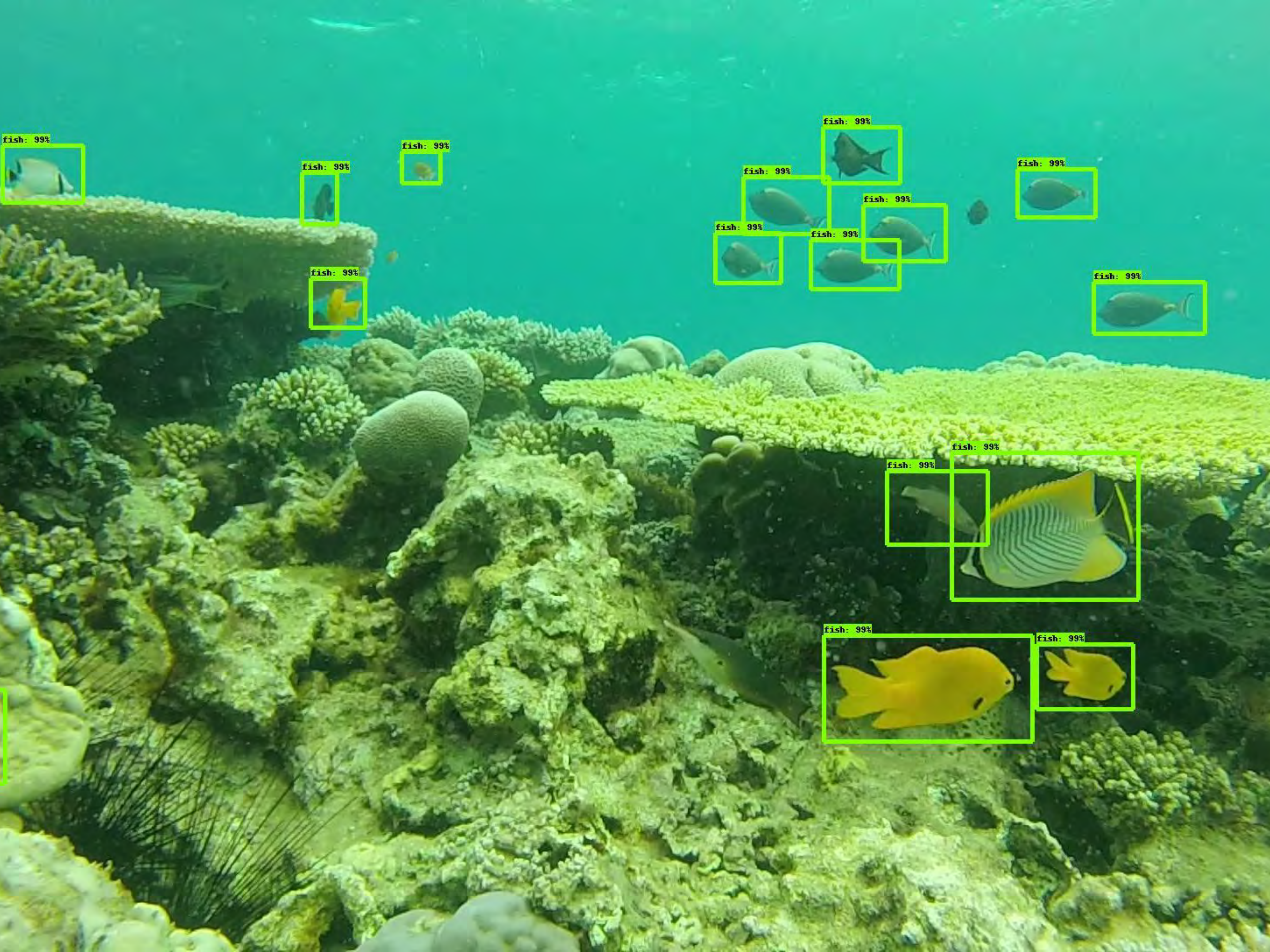
Poisson/Pas poisson via Faster-RCNN

- Apprentissage :
  - Méditerranée : 7 vidéos, 300 frames, 1268 vignettes.
  - Mayotte: 5 vidéos, 110 frames, 866 vignettes.
- Test:
  - Mayotte : 1 vidéo, 23 frames, entre 18 et 25 individus par frames



- Rappel moyen: 0.78
  (on rate quelques poissons)
- Précision moyenne : 0.99
  (mais on en invente aucun)

Ren S., He K., Girshick R., Sun J. " Faster r-cnn: Towards real-time object detection with region proposal networks ". In NIPS'2015
Zoph B., Vasudevan V., Shlens J., Le Q. V. Learning transferable architectures for scalable image recognition. In CVPR'2018
Liu C., Zoph B., Neumann M., Shlens J., Hua W., Li L. J., ... Murphy, K. " Progressive neural architecture search. In ECCV'2018 -> PNAS

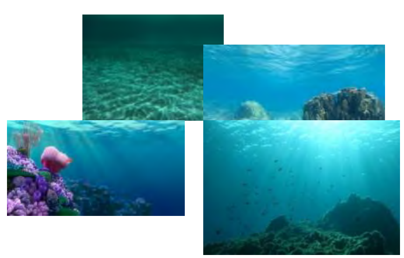

fish: 99%
fish: 99%
fish: 99%
fish: 99%
fish: 99%
fish: 99%
fish: 99%
fish: 99%
fish: 99%
fish: 99%
fish: 99%
fish: 99%
fish: 99%
fish: 99%
fish: 99%
fish: 99%

Classification automatique
à partir d'imagettes

Utilisation du Deep Learning
(réseau de neurones convolutionnel)

– Apprentissage via une grandes bases d'images (+ de 121 espèces pour le moment / objectif de ~ 300 espèces)

Species	Thumbnails
Abudefduf sparoides	2482
Abudefduf vaigiensis	11328
Chaetodon trifascialis	2912
Chromis weberi	7152
Dascyllus carneus	4552
Lutjanus kasmira	3300
Monotaxis grandoculis juvenile	2478
Mulloidichthys vanicolensis	3764
Myripristis botche	2528
Naso elegans	4138
Naso vlamingii	3578
Nemateleotris magnifica	2372
Odonus niger	5972
Plectroglyphidodon lacrymatus	1304
Pomacentrus sulfureus	10352
Pseudanthias squamipinnis male	2946
Pygoplites diacanthus	2212
Thalassoma hardwicke	3158
Zanclus cornutus	3772
Zebrasoma scopas	3670
Background	862174
Part of fish	512555



Species
Acanthurus lineatus
Acanthurus nigrofuscus
Chromis ternatensis
Chromis viridis/Chromis atripectoralis
Pomacentrus sulfureus
Pseudanthias squamipinnis
Zebrasoma scopas
Ctenochatus striatus
Random/specific background
Part of Fish

Humain vs Machine
tache de classification d'imagettes

Species	Thumbnails
Abudefduf sparoides	88
Abudefduf vaigiensis	47
Chaetodon trifascialis	149
Naso elegans	165
Pomacentrus sulfureus	443
Pygoplites diacanthus	35
Thalassoma hardwicke	73
Zanclus cornutus	53
Zebrasoma scopas	144
Total	1197

Images testées

Species	Network results	Human Results
Abudefduf sparoides	93.4	87.73
Abudefduf vaigiensis	97.3	84.68
Chaetodon trifascialis	95.1	89.42
Naso elegans	98.4	94.81
Pomacentrus sulfureus	97.9	93.23
Pygoplites diacanthus	90.4	77.38
Thalassoma hardwicke	96	91.01
Zanclus cornutus	97.1	97.82
Zebrasoma scopas	96.2	88.26
Moyenne	95.7	89.3

Taux de bonne classification

- Temps idenfication :
  - Expert Humain (11) : **~ 5 secondes**
  - Machine : **~ 0.06 secondes** sur notre plateforme

Images : 128x128x3.
Humain : test dure ~ 20 minutes et un expert teste ~ 270 images ;
Machine : GPU Nvidia Titan X ~ 3000 coeurs; apprentissage sur 1,100,000 vignettes, 70 epochs - 14 jours

Les erreurs

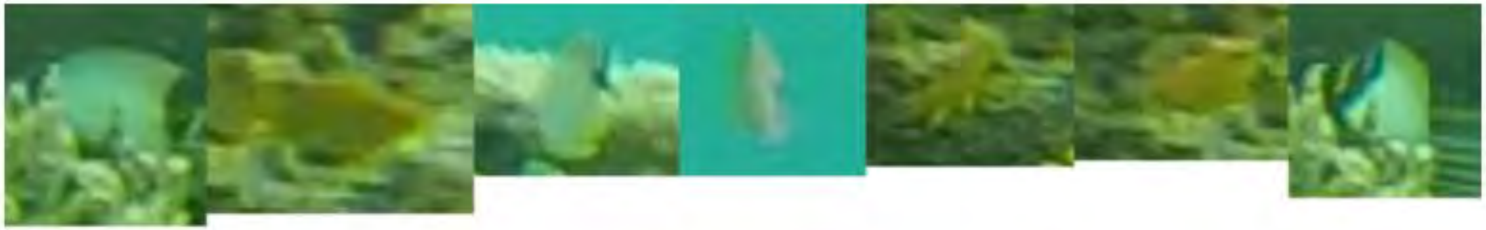


Figure 5: Sample of pictures recognized by the network and not recognized by experts.

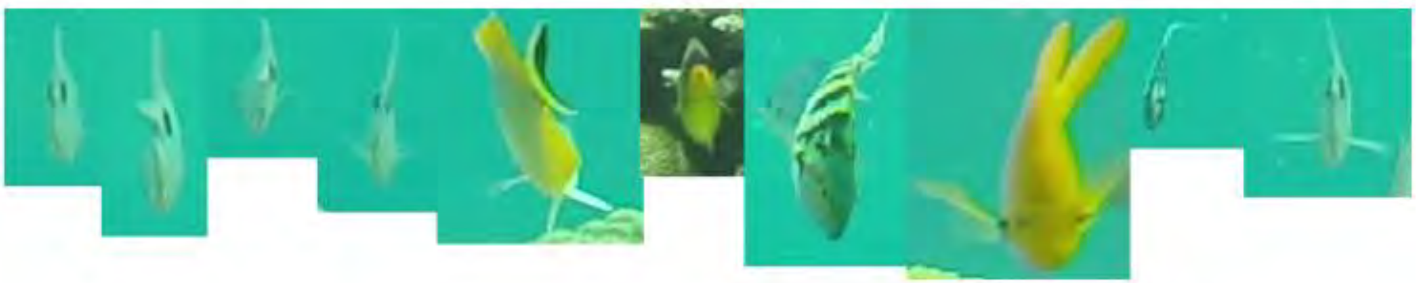
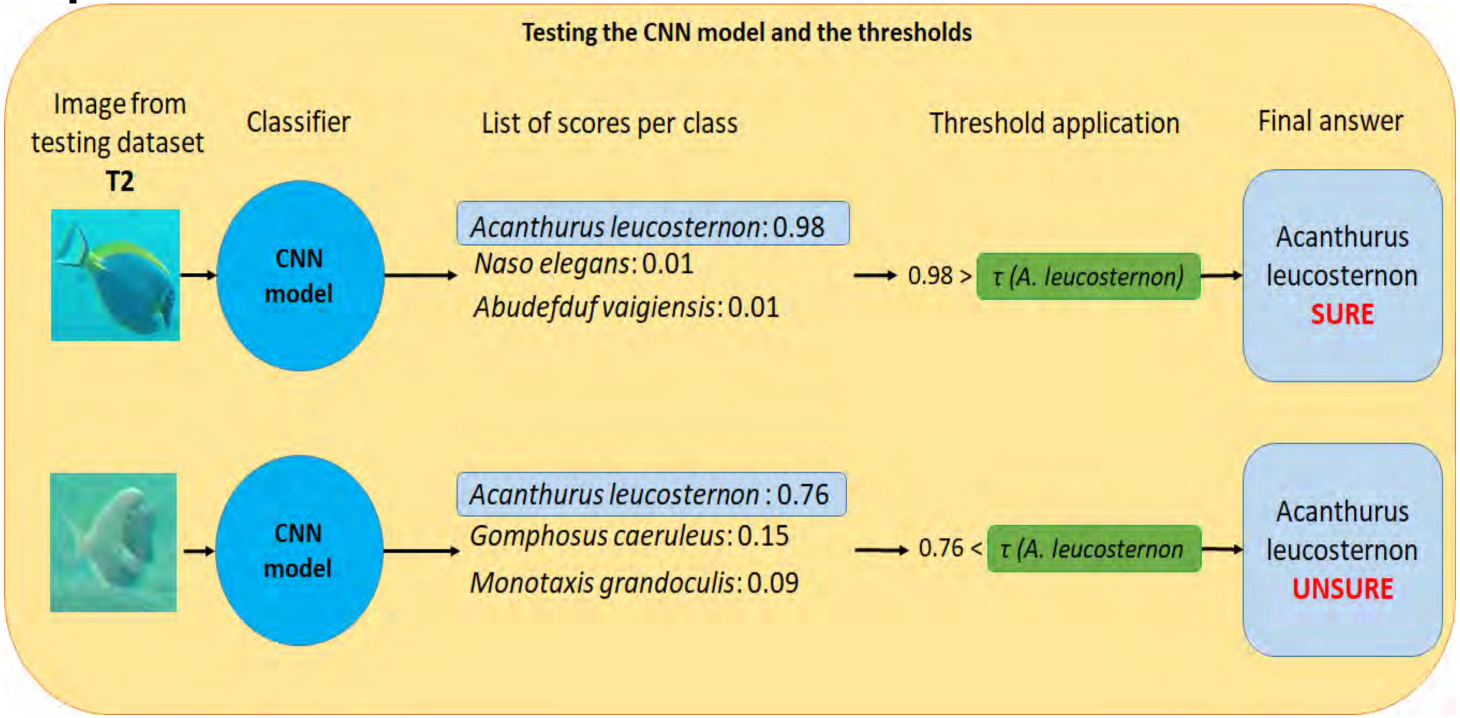


Figure 6: Samples of pictures recognized by experts and not recognized by the network.

Figure 7: Sample of pictures misclassified by both humans and network.

Gestion « sûr » / pas « sûr »

- **Post-processing:**
  Définition d'une classe « le réseau n'est pas sûr » en plus des classes contenant les espèces



Choix d'un scénario

Par espèce :

- Seuil 1 : meilleur classif

  $$\min \left[\underbrace{\#\{\text{Misclassif}\}}_{\textit{taux mauvaise classif.}} \mid (\#\{\underbrace{\text{Vrai Positif}\} = \text{maximum})}_{\textit{taux classification correcte}} \right]$$

- Seuil 2 : borne l'erreur de classif

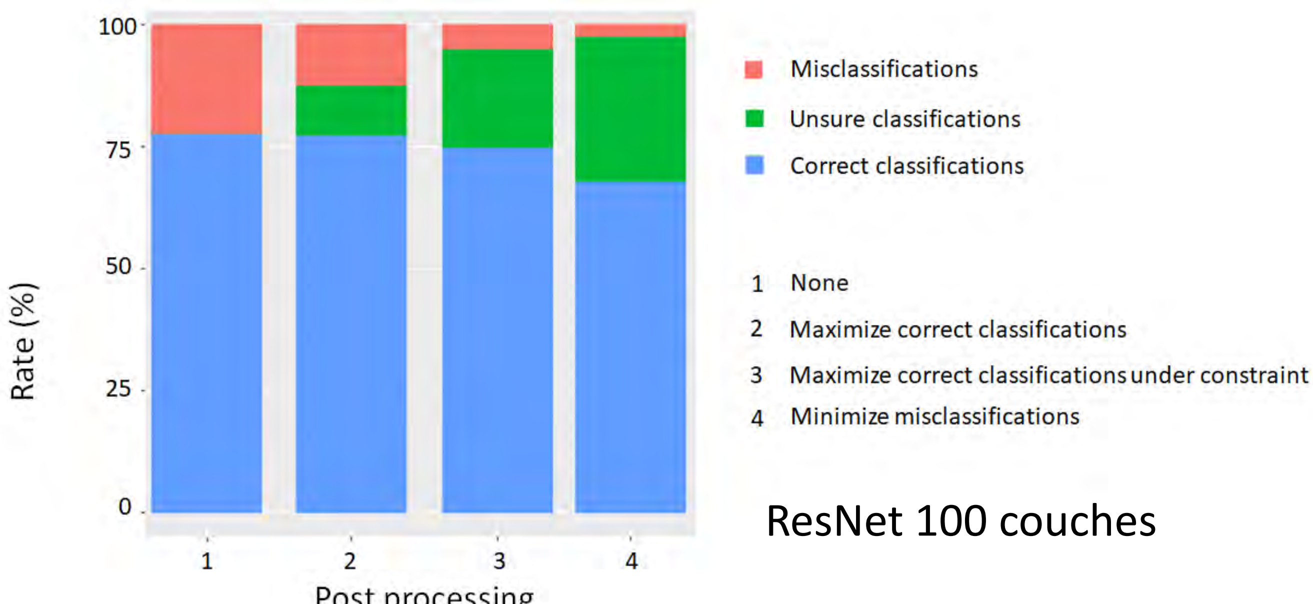
  $$\max \left[\underbrace{\#\{\text{Vrai Positif}\}}_{\textit{taux classif. correcte}} \mid \underbrace{(\#\{\text{Misclassif}\} < 5\ \%)}_{\textit{taux mauvaise classification}} \right]$$

- Seuil 3 : erreur de classif minimum

  $$\max \left[\underbrace{\#\{\text{Vrai Positif}\}}_{\textit{taux classif. correcte}} \mid \underbrace{(\#\{\text{Misclassif}\} = \text{minimum})}_{\textit{taux mauvaise classification}} \right]$$

Résultats

- Data-set (Ile de Mayotte) :
  - D0 : 130 vidéos, 69 169 images, 20 espèces. Entre 1 134 et 7 345 imagettes par espèce.
  - D1 : 20 vidéos, 6 320 images
  - D2 : 25 vidéos, 13 232 images



ResNet 100 couches

Les scénarios…

Qui va traiter les « unsures » ?

- CAS1 et CAS2 : Personne

  Détection d’événements (peu importe le taux d’erreur)

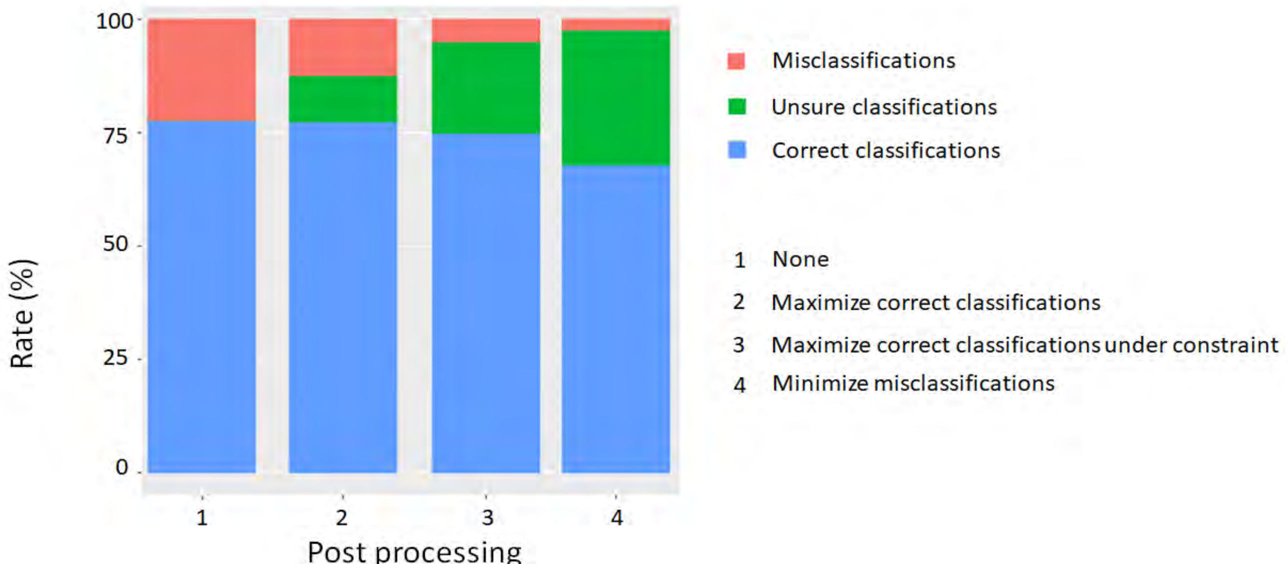
  - monitoring d’espèces invasives
  - évenement rare

- CAS3: Des humains ; les “unsures” sont <u>vraiment</u> “unsure”

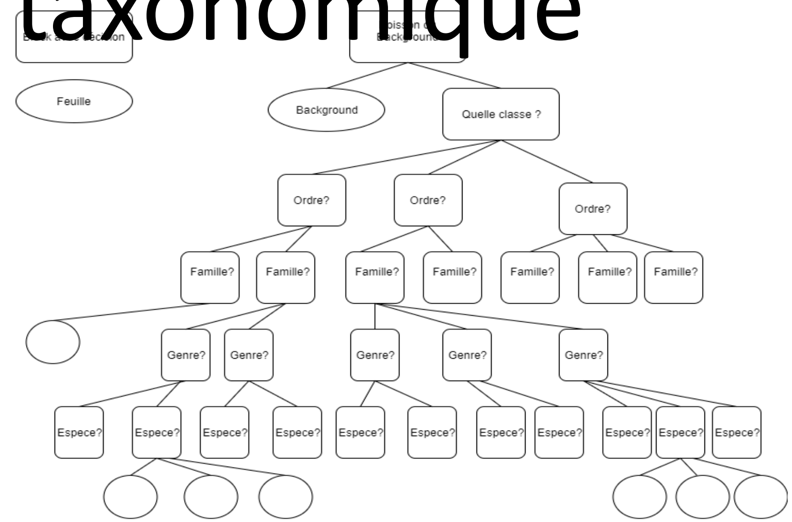
  Gain de temps pour le traitement de gros volumes

- CAS4: L’expert

  L’expert recherche une qualité d’annotation sur un petit data-set



Utilisation de l’arbre taxonomique



Post-traitement par rapport à l’arbre taxonomique :

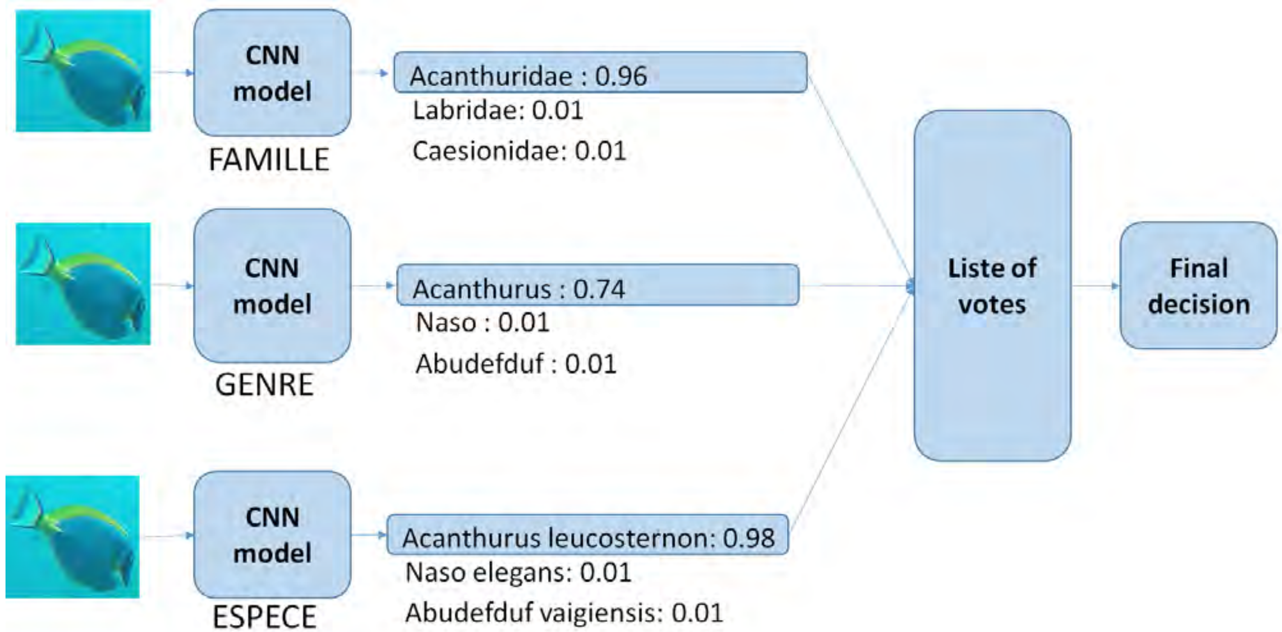
CAS1: Les votes (confiants) sont cohérents

On classe au plus fin (espèce/genre/famille)

CAS2: Les votes (confiants) sont incohérents

2 d’accords -> la majorité l'emporte

Non d’accords -> classé dans « Unsure ».



- **Règne** (*Regnum*)
- Sous-règne (*Subregnum*)
- Rameau (*Ramus*, « branch » en anglais)
- Infra-règne (*Infraregnum*)
  - Super-embranchement, Super-division (*Superphylum*, *Superdivisio*)
  - **Embranchement**, **Division** (***Phylum***, ***Divisio***)[2]
  - Sous-embranchement, Sous-division (*Subphylum*, *Subdivisio*)
  - Infra-embranchement (*Infraphylum*)
  - Micro-embranchement (*Microphylum*)
    - Super-classe (*Superclassis*)
    - **Classe** (***Classis***)
    - Sous-classe (*Subclassis*)
    - Infra-classe (*Infraclassis*)
      - **Super-ordre** (*Superordo*)
      - **Ordre** (*Ordo*)
      - Sous-ordre (*Subordo*)
      - Infra-ordre (*Infraordo*)
      - Micro-ordre (*Microordo*)
        - Super-famille (*Superfamilia*)
        - **Famille** (***Familia***)
        - Sous-famille (*Subfamilia*)
        - Tribu (*Tribus*)
        - Sous-tribu (*Subtribus*)
          - **Genre** (***Genus***)
          - Sous-genre (*Subgenus*)
          - Section (*Sectio*)

-> Supprime les erreurs quand incohérences.

-> Permet de réduire les classifications incorrectes.

SeaCLEF 2017

The CLEF Association is an independent no-profit legal entity, established in October 2013.

- Tâche 4: Reconnaissance individuelle des baleines
- Analyse des nageoires caudales

Objectif : A partir d'une image de caudale, trouver l'individu correspondant dans la base

Aire de Madagascar

C'est le jeu des 7 différences !

Unsupervised identification

Alexis Joly, Jean-Christophe Lombardo, Julien Champ, Anjara Saloma

https://www.imageclef.org/lifeclef/2017/sea

Bonnes correspondances (très peu)

Mauvaises correponcances (beaucoup)

20130909_1744_G1A_TF2

20140812_1978_G7inconnu_TF2

20140727_1900_G5A_TF1

150809_KA_MNG2A_T1

Une solution = traitement du signal + recherche dans une base

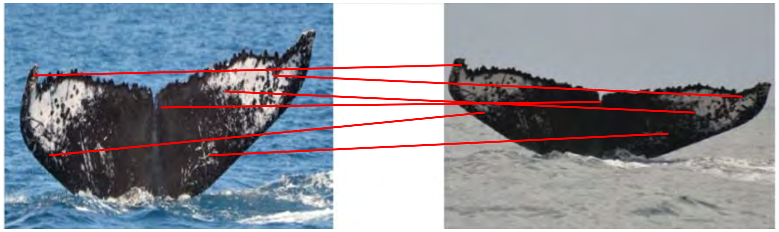
La cohérence spatial (répartition spatiale) des marqueurs biologique locaux est une information crucial pour réduire les faux positifs.

- Principe :

  1) **Trouver des zones** caractéristiques (points) de l'images (traitement du signal; SIFT)

  2) **Appareiller les points** et « filtrer »
  Estimer une transformation géométrique entre l'image à appareiller et l'image cible pour affiner l'apparition

  3) **Evaluer la "similitude"** entre les deux individus comparés via une métrique de « scoring » des zones appareillés.



SIFT-based matching

ConvNet features

Run name	Average Precision
ZenithINRIA_SiftGeo	0,49
ZenithINRIA_SiftGeo_QueryExpansion	0,43
ZenithINRIA_GoogleNet_3layers_borda	0,33
bmetmit_whalerun_1	0,25
bmetmit_whalerun_3	0,10

En 1D la direction de la pente est égale au négatif de la dérivé ☺

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

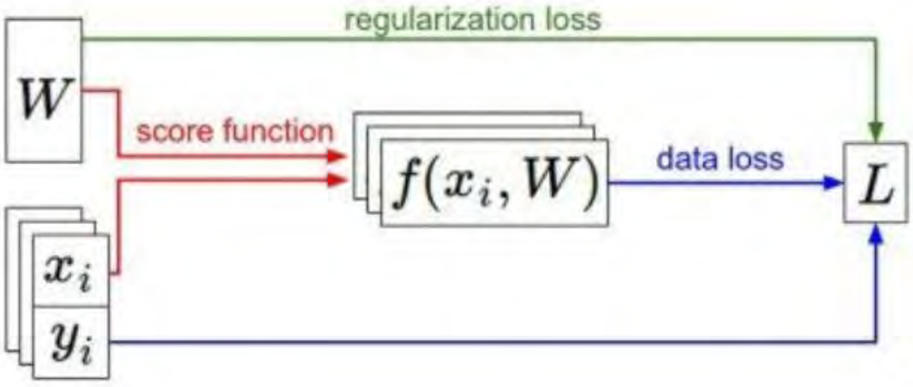
Softmax

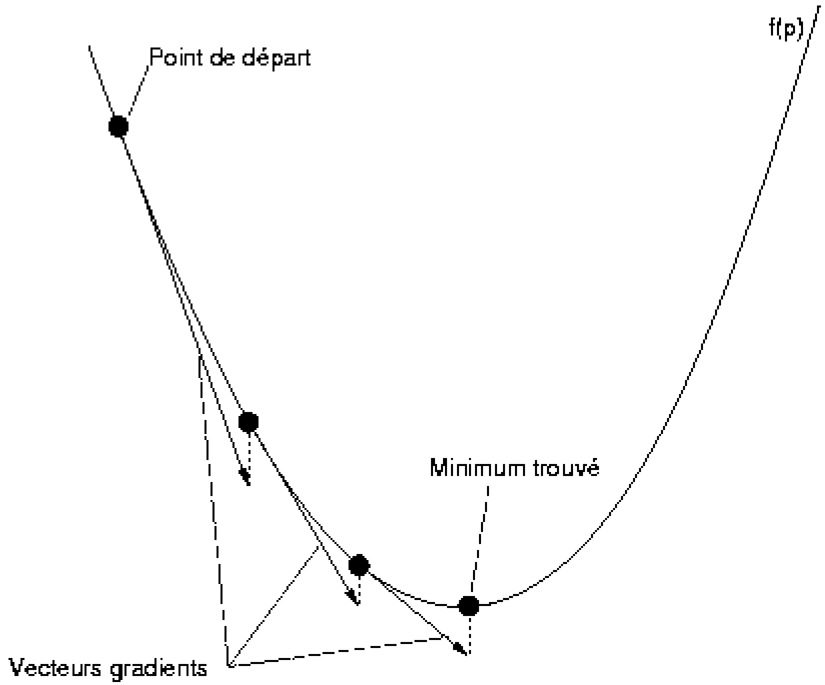
$$L_i = -\log\left(\frac{e^{s_{y_i}}}{\sum_j e^{s_j}}\right)$$

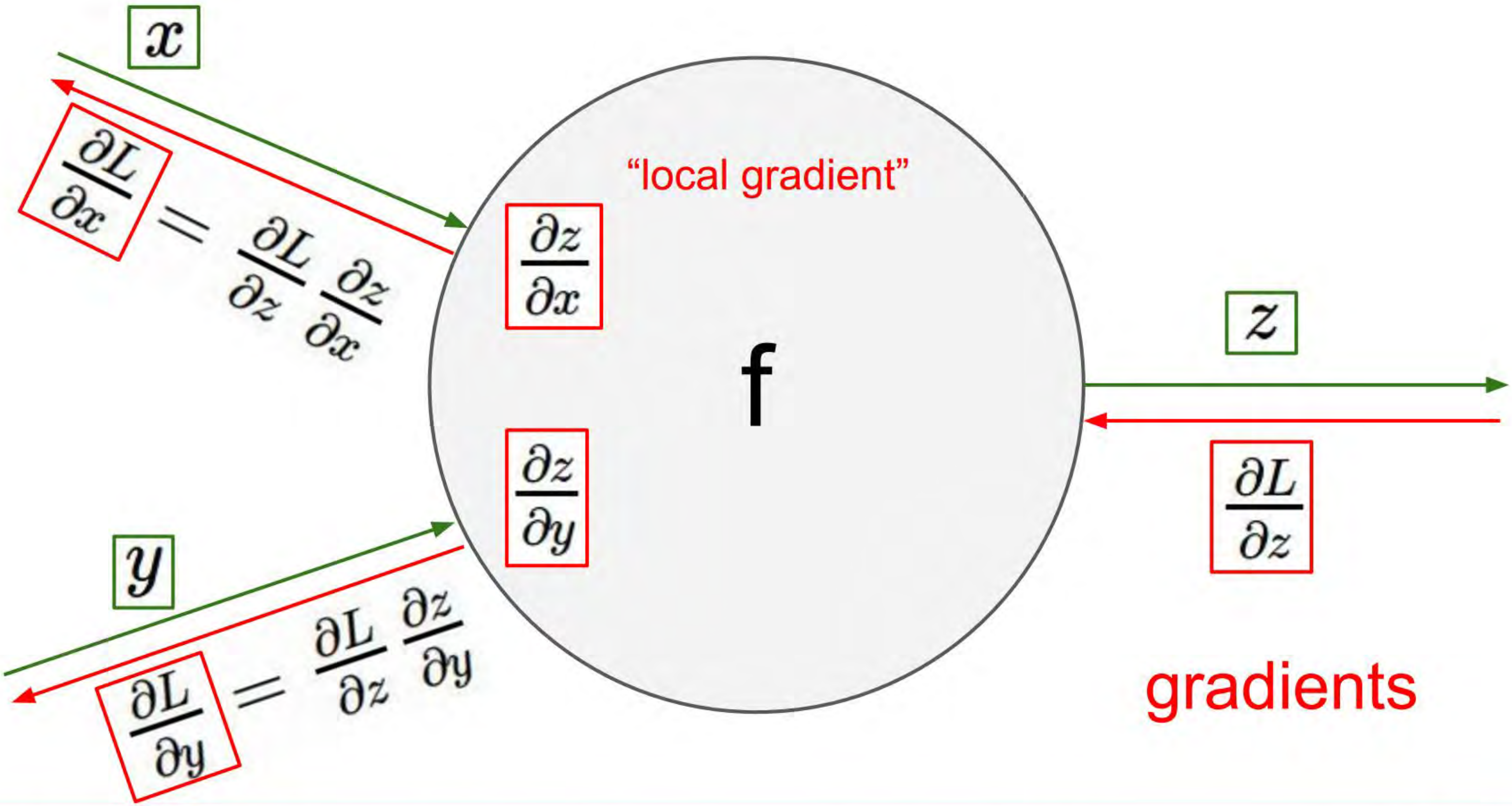
SVM

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

$$L = \frac{1}{N} \sum_{i=1}^{N} L_i + R(W)$$ Full loss






x
$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial x}$
"local gradient"
$\frac{\partial z}{\partial x}$
f
z
$\frac{\partial z}{\partial y}$
$\frac{\partial L}{\partial z}$
y
$\frac{\partial L}{\partial y} = \frac{\partial L}{\partial z}\frac{\partial z}{\partial y}$
gradients

SGD + Momentum

SGD

$$x_{t+1} = x_t - \alpha \nabla f(x_t)$$

```
while True:
  dx = compute_gradient(x)
  x += learning_rate * dx
```

SGD+Momentum

$$v_{t+1} = \rho v_t + \nabla f(x_t)$$
$$x_{t+1} = x_t - \alpha v_{t+1}$$

```
vx = 0
while True:
  dx = compute_gradient(x)
  vx = rho * vx + dx
  x += learning_rate * vx
```

- Build up "velocity" as a running mean of gradients
- Rho gives "friction"; typically rho=0.9 or 0.99

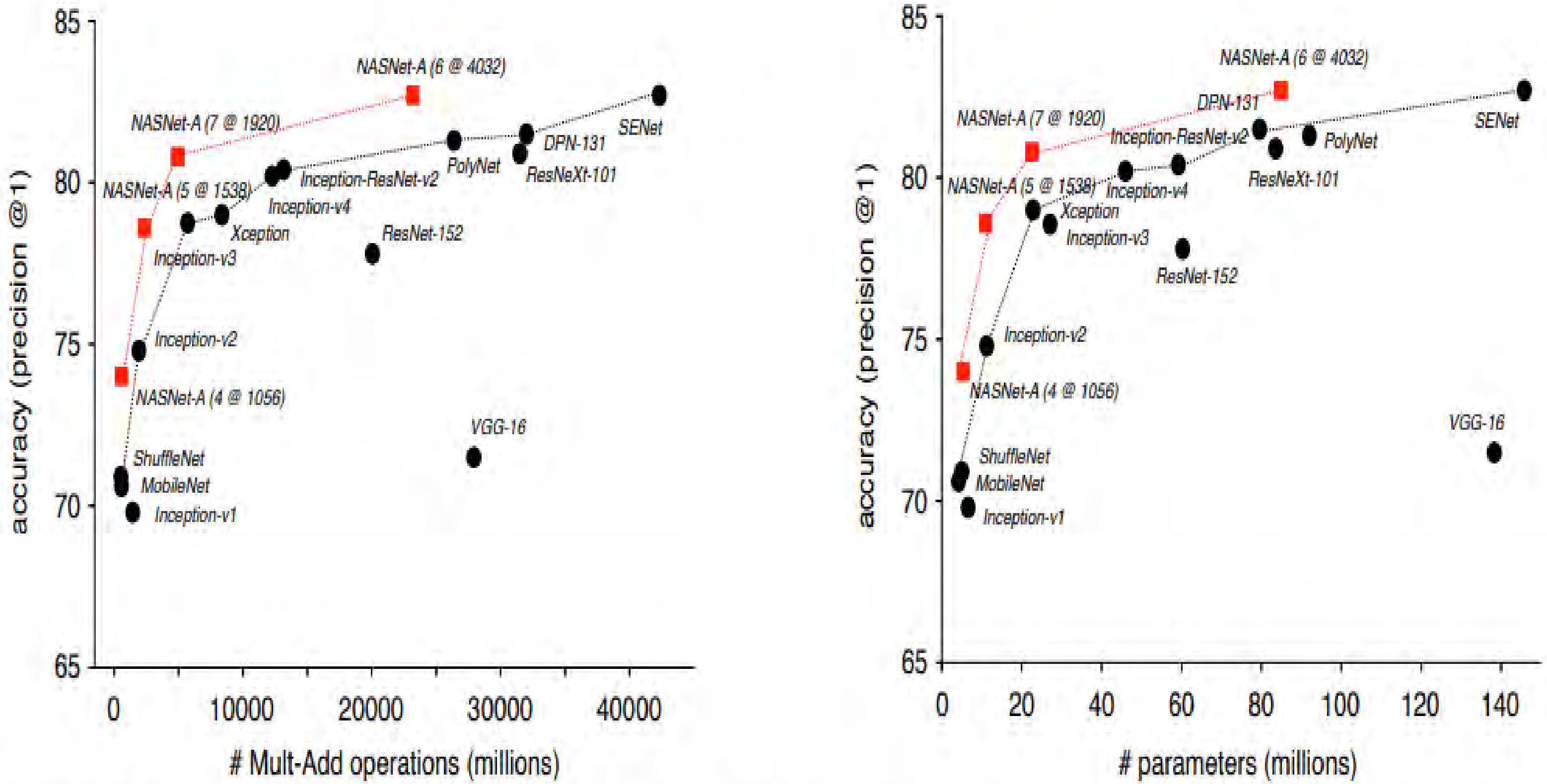


Figure 5. Accuracy versus computational demand (left) and number of parameters (right) across top performing published CNN architectures on ImageNet 2012 ILSVRC challenge prediction task. Computational demand is measured in the number of floating-point multiply-add operations to process a single image. Black circles indicate previously published results and red squares highlight our proposed models.