



HAL
open science

Deep-Learning for the observation of marine/terrestrial life. A case study: the localization and identification of fish species

Marc Chaumont

► To cite this version:

Marc Chaumont. Deep-Learning for the observation of marine/terrestrial life. A case study: the localization and identification of fish species. 2019. lirmm-03820892

HAL Id: lirmm-03820892

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-03820892>

Submitted on 19 Oct 2022

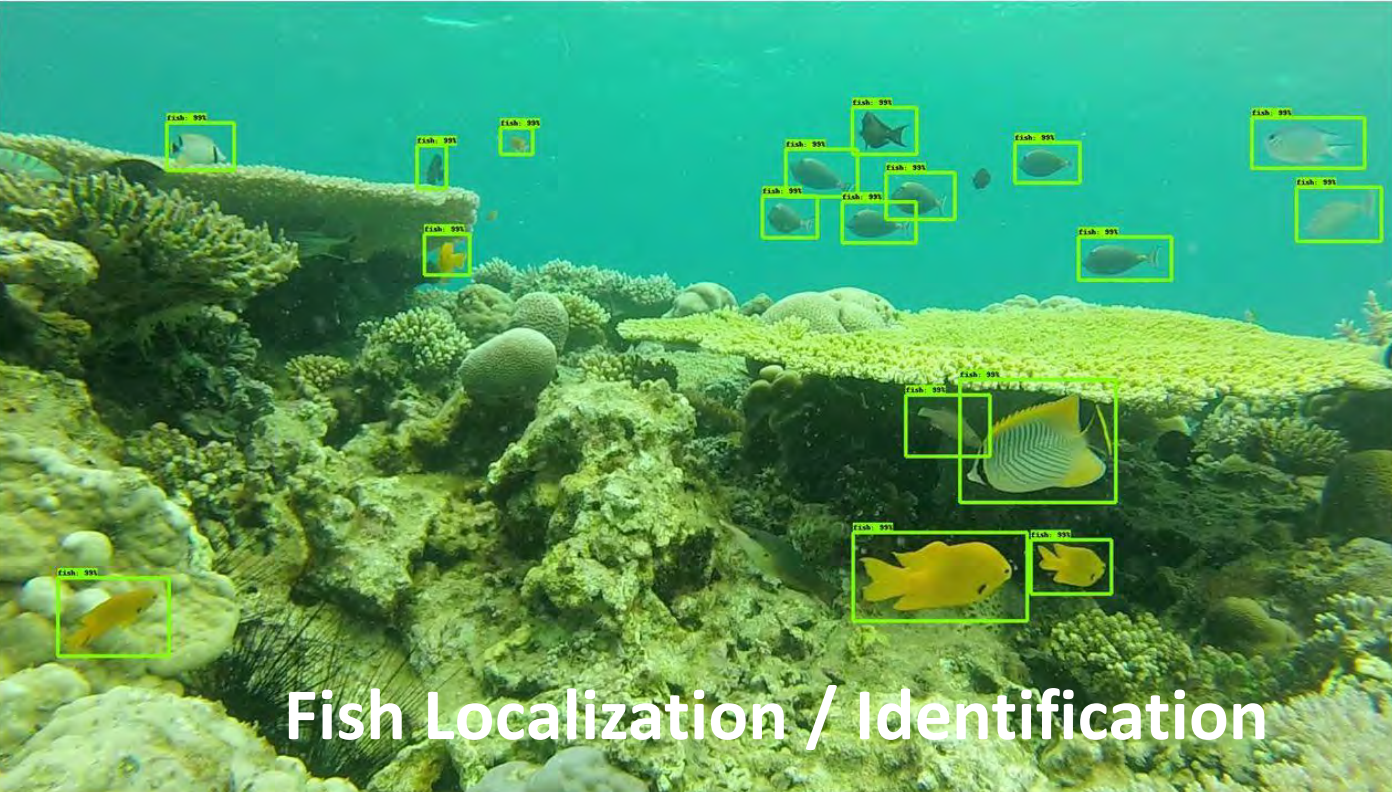
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

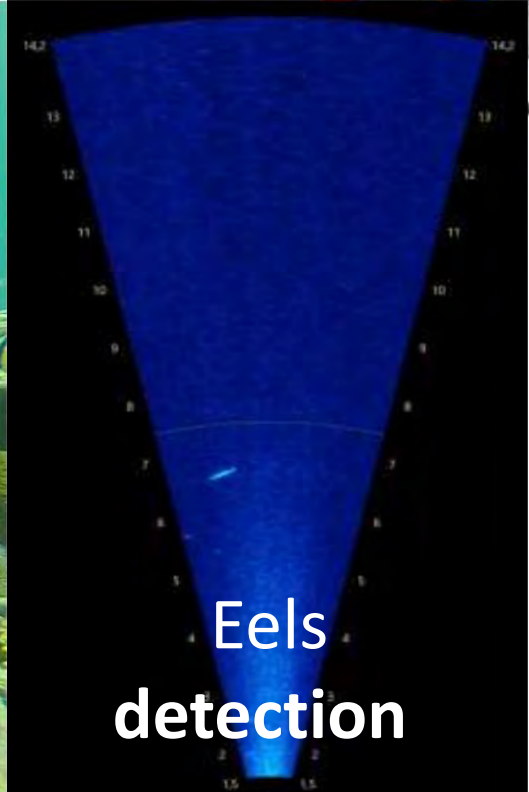
Deep-Learning for the observation of marine/terrestrial life

Marc Chaumont

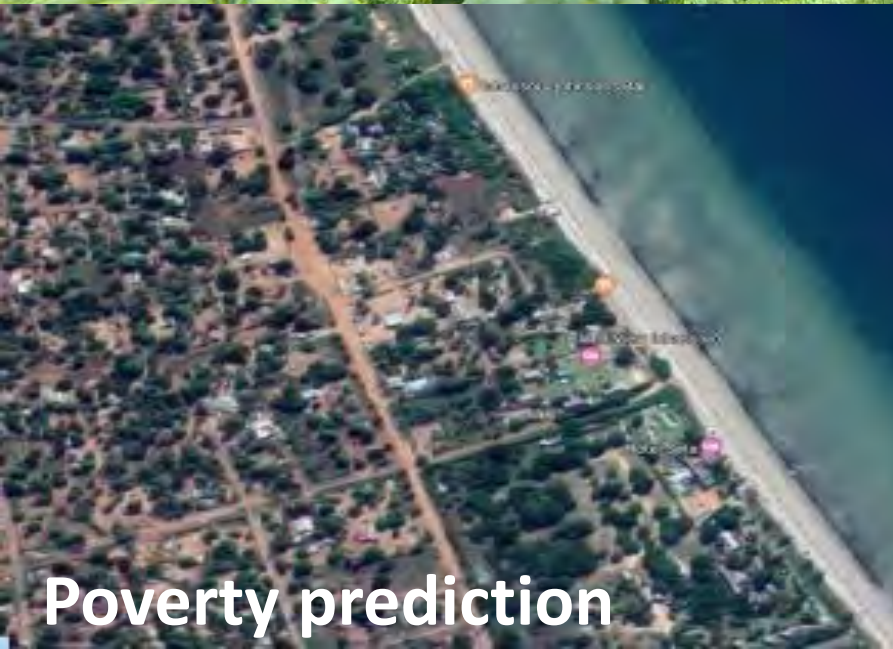




Fish Localization / Identification



Eels detection



Poverty prediction



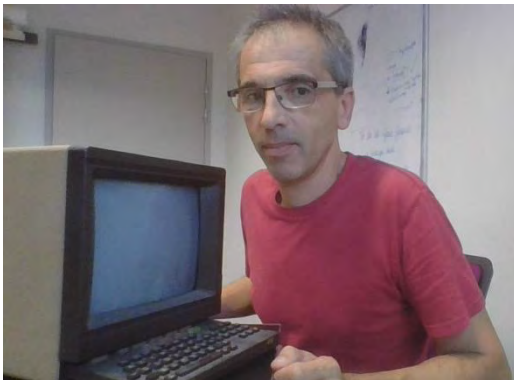
Shark localization
under DCP
with multiple
cameras

<http://www.peche.pf>



Species classifications

LIRMM researchers involved



Gérard Subsol

Marc Chaumont

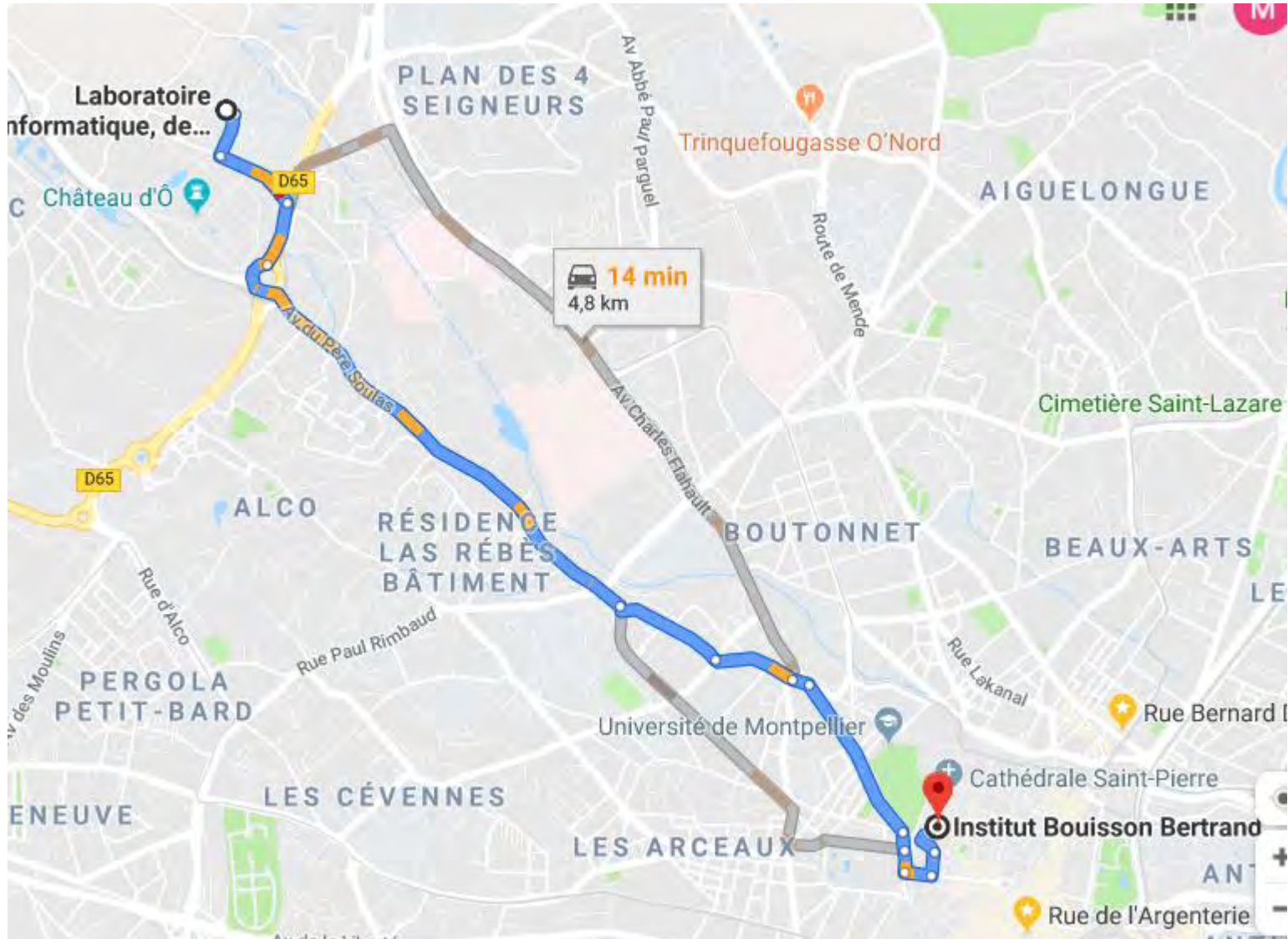
Working on several applications of 2D and 3D image processing.

Steganography and more generally image and video processing

Computer science, image processing, modelling

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

LIRMM (14 min from here)





LIRMM

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

Buildings



Historical building



New building

Outline

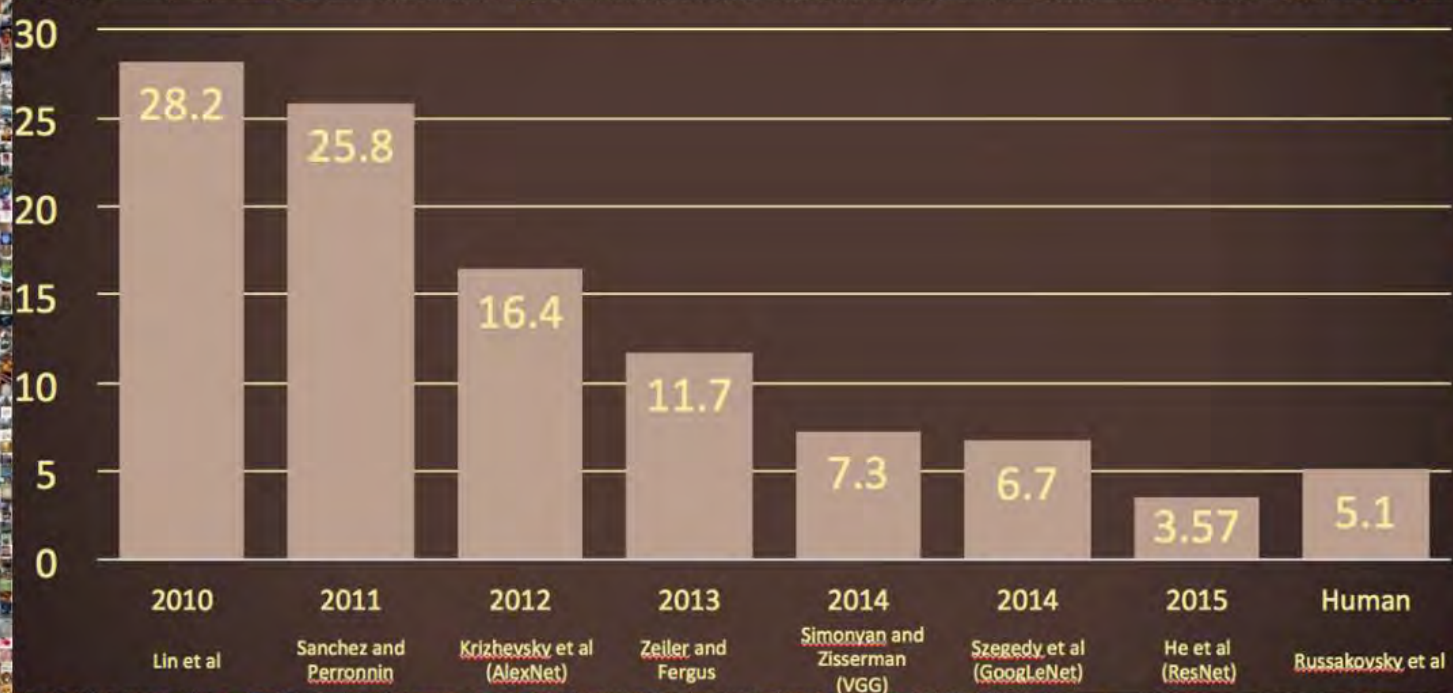
- Few words on Deep Learning
- Few projects done in our « subset »-team
(G rard and me)

IMAGENET Large Scale Visual Recognition Challenge

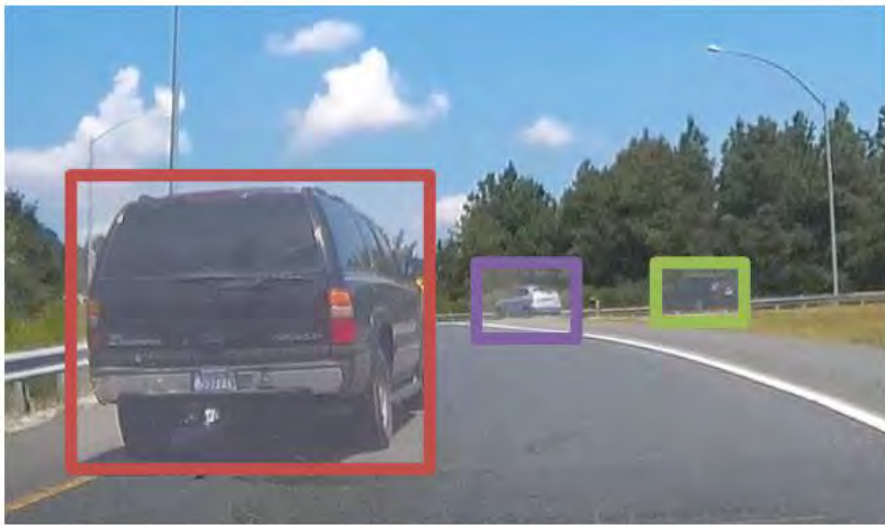
The Image Classification Challenge:

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

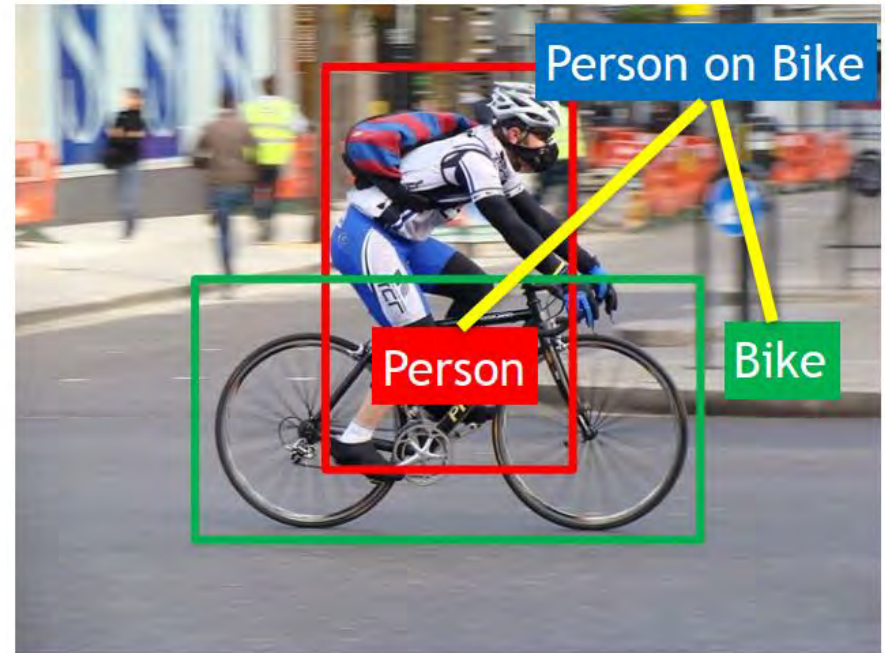
- Object detection
- Action classification
- Image captioning
- ...



Person

Hammer

This image is licensed under [CC BY-SA 2.0](#); changes made

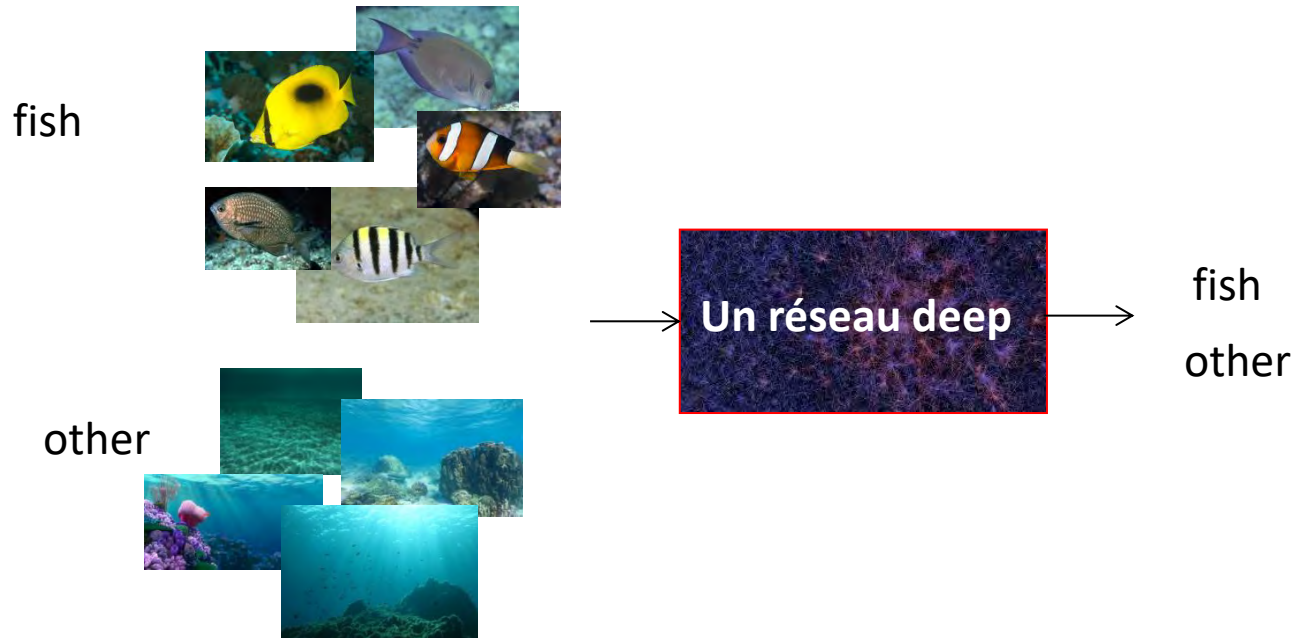


This image is licensed under [CC BY-SA 3.0](#); changes made

The learning protocol (supervised case)

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

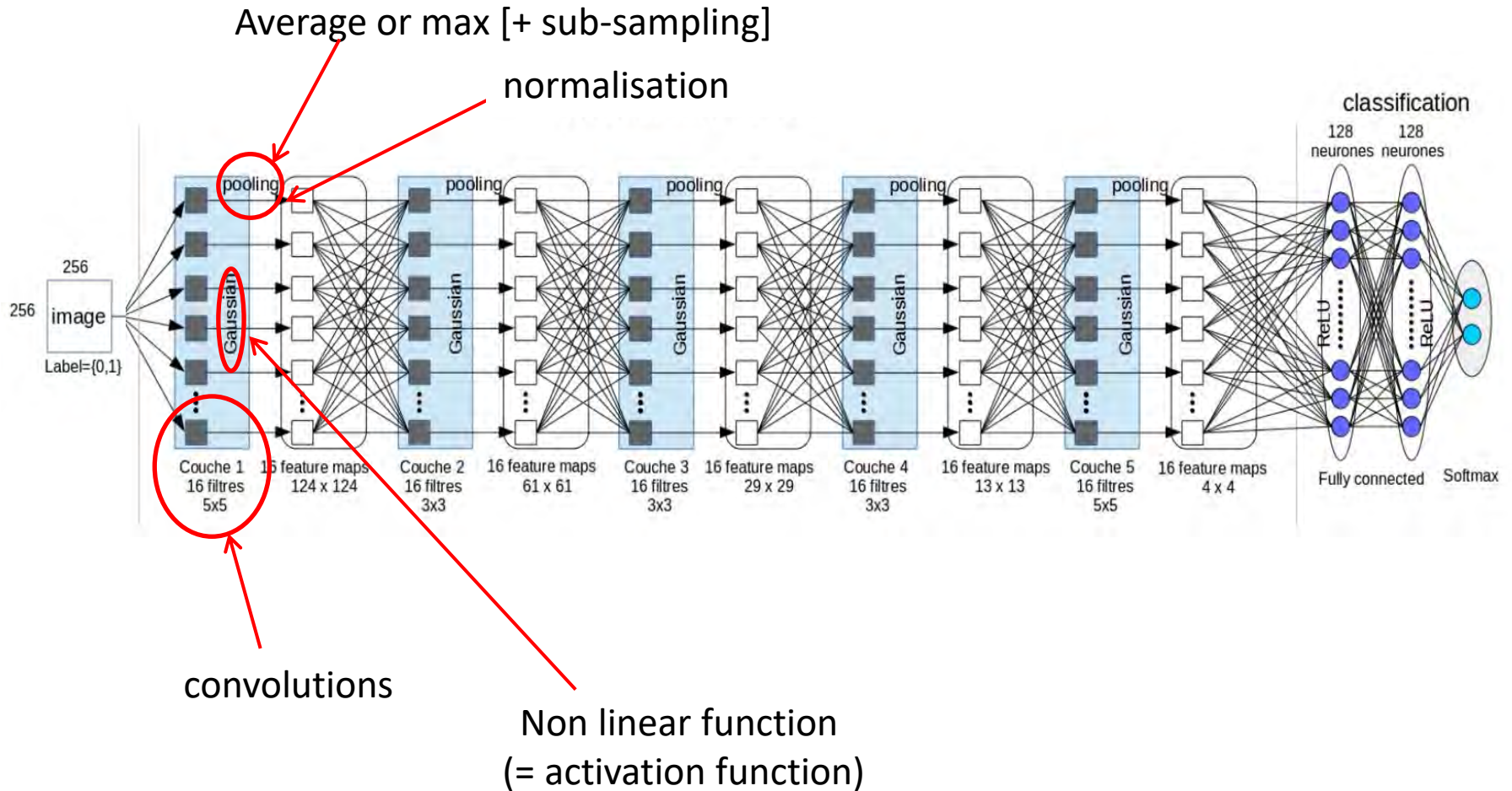
THE LEARNING



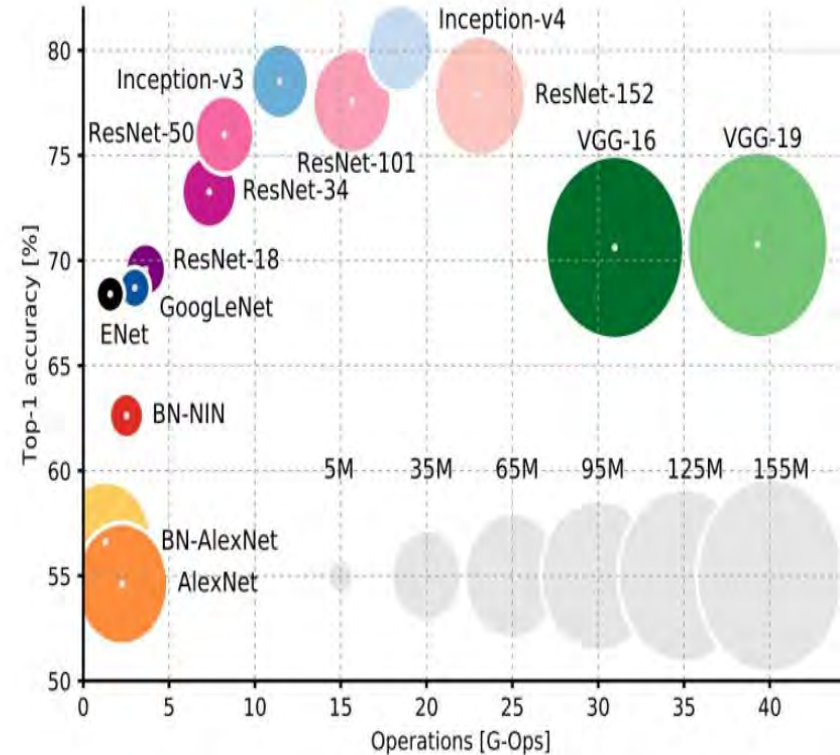
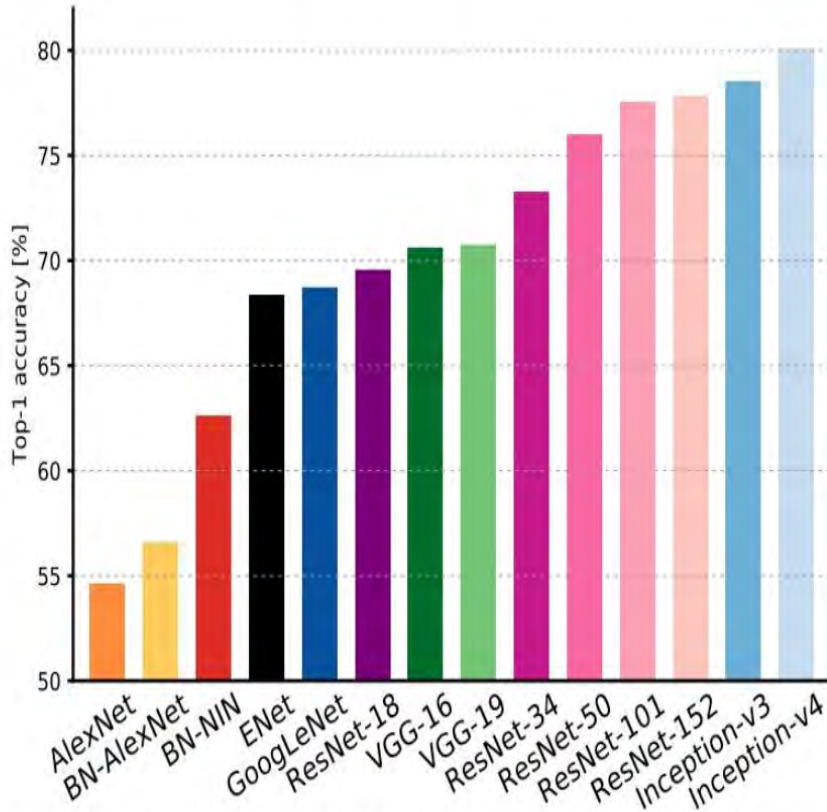
- STEP 2) We use the network 😊

A CNN

Convolutional Neural Network



Comparing complexity...

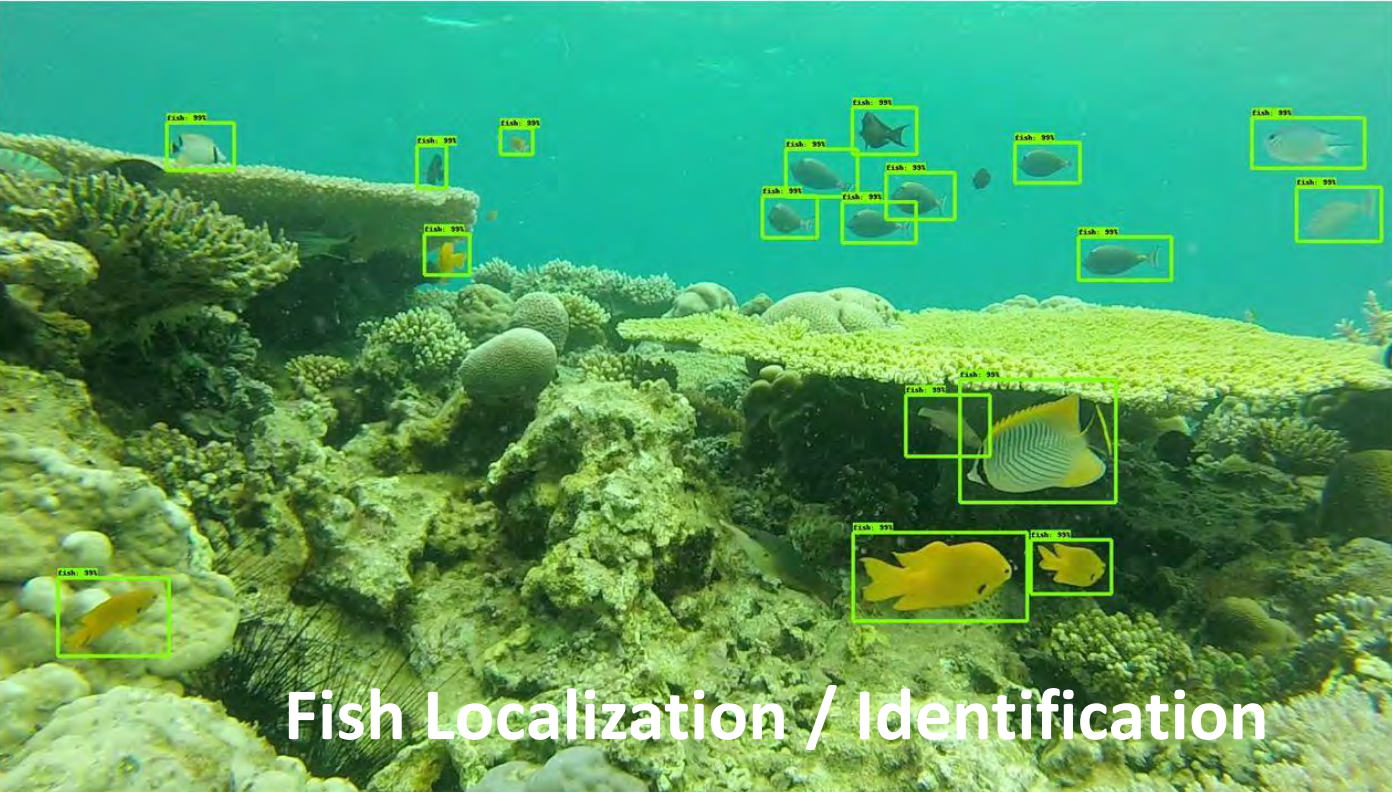


An Analysis of Deep Neural Network Models for Practical Applications, 2017.

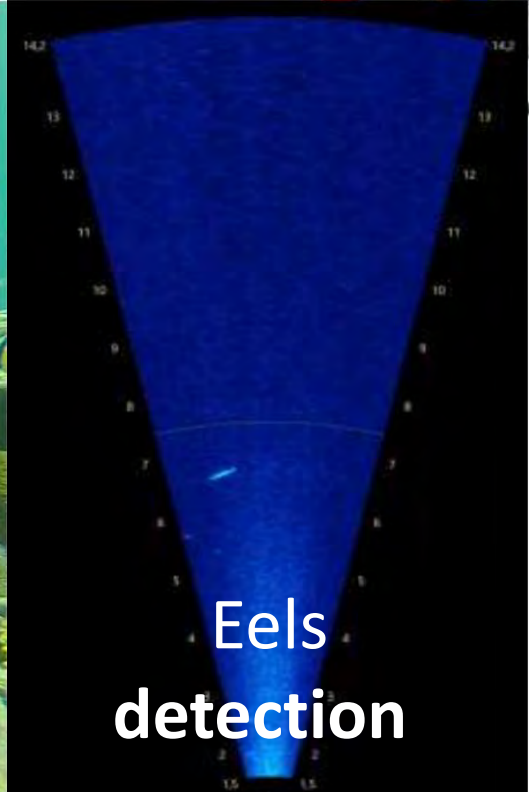
Figures copyright Alfredo Canziani, Adam Paszke, Eugenio Culurciello, 2017. Reproduced with permission.

Outline

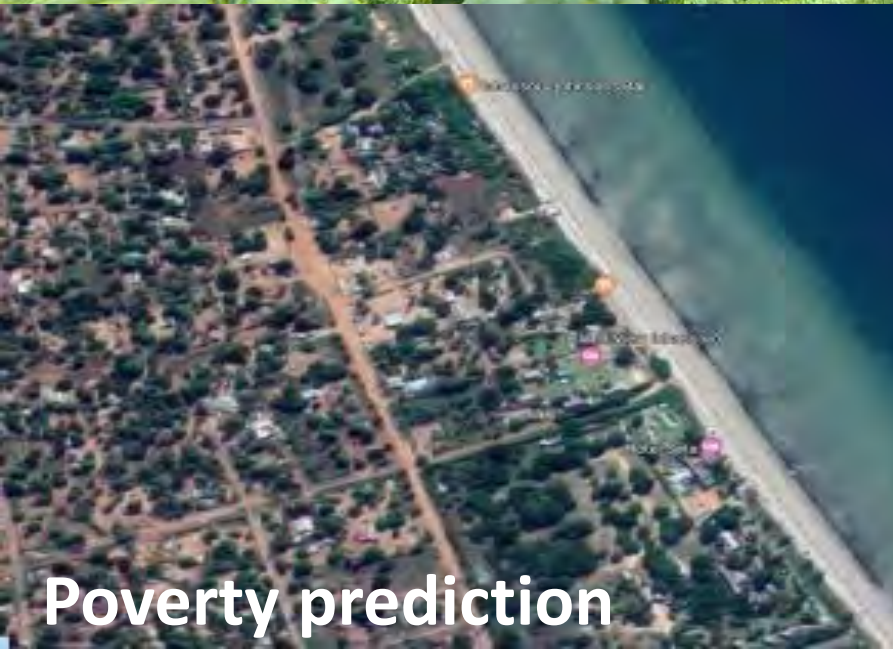
- Few words on Deep Learning
- Few projects done in our « subset »-team
(G rard and me)



Fish Localization / Identification



Eels detection

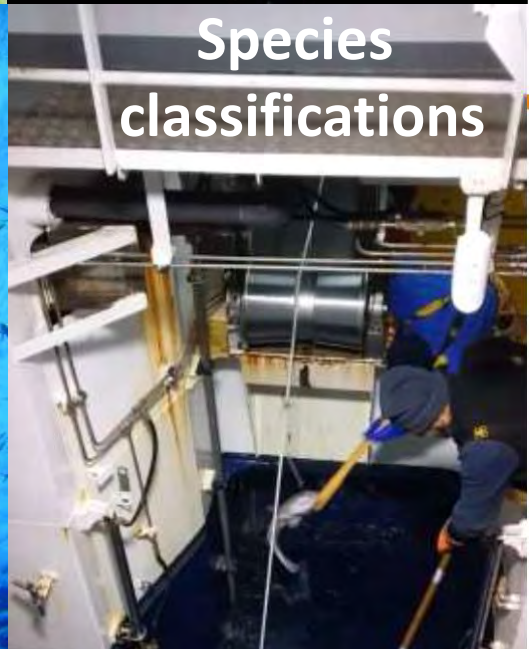


Poverty prediction



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,

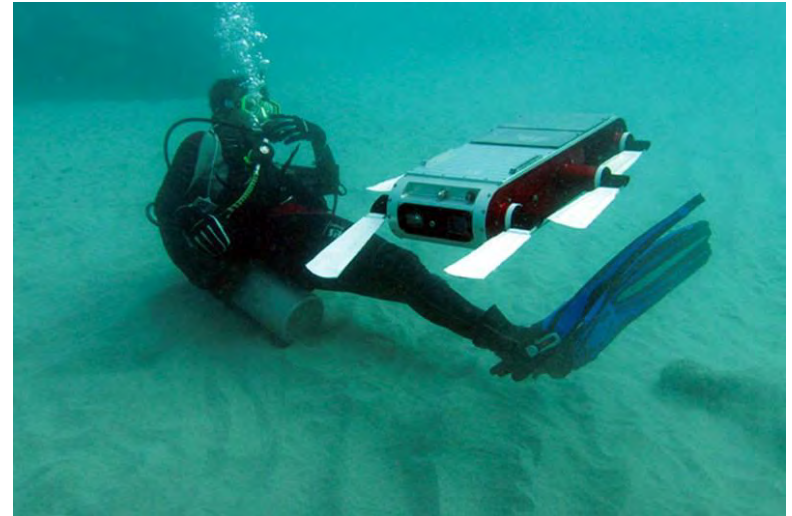


Gérard Subsol

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

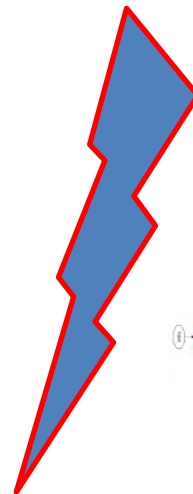
The Machine



20 minutes /
Human



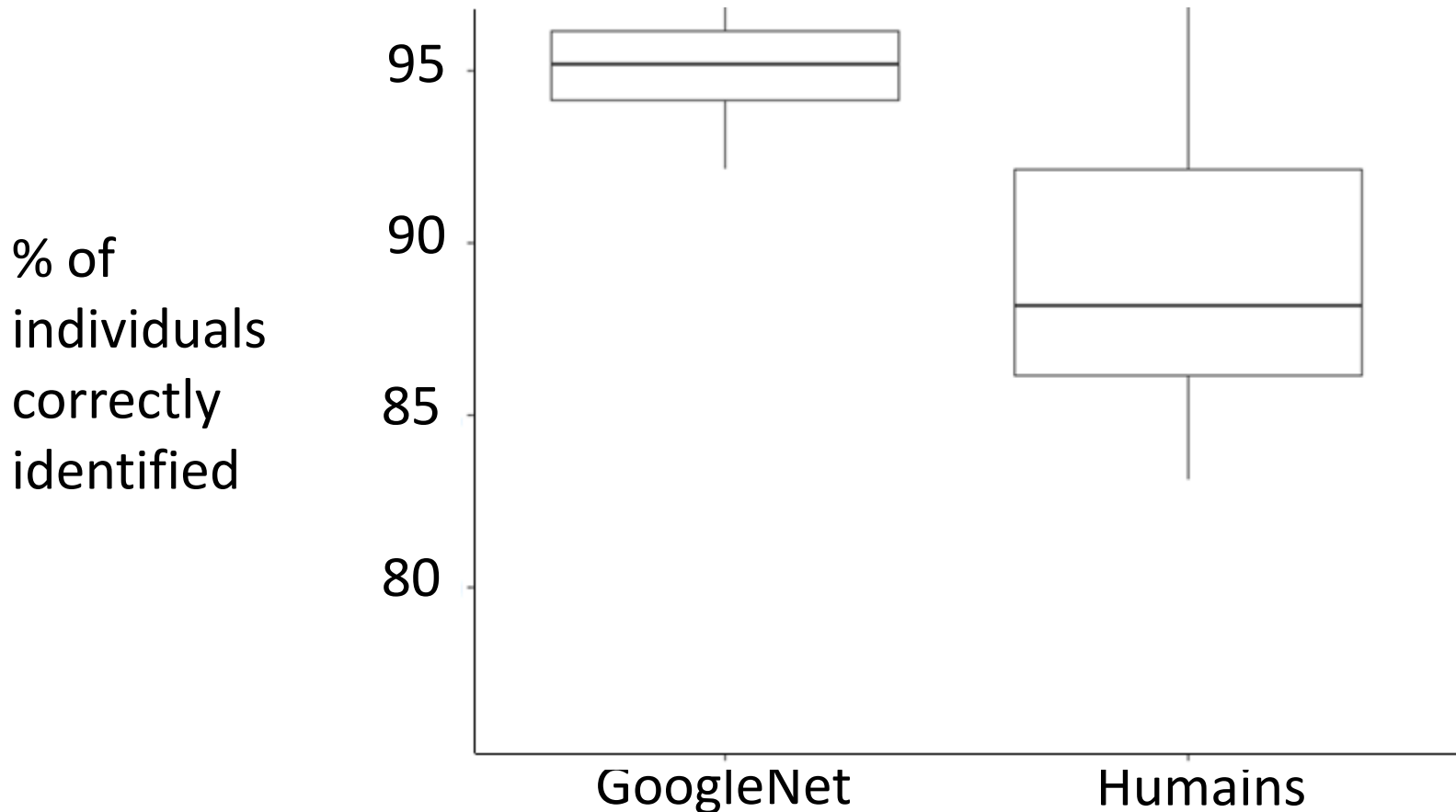
14 Humans



GoogLeNet
(trained on 20 species)



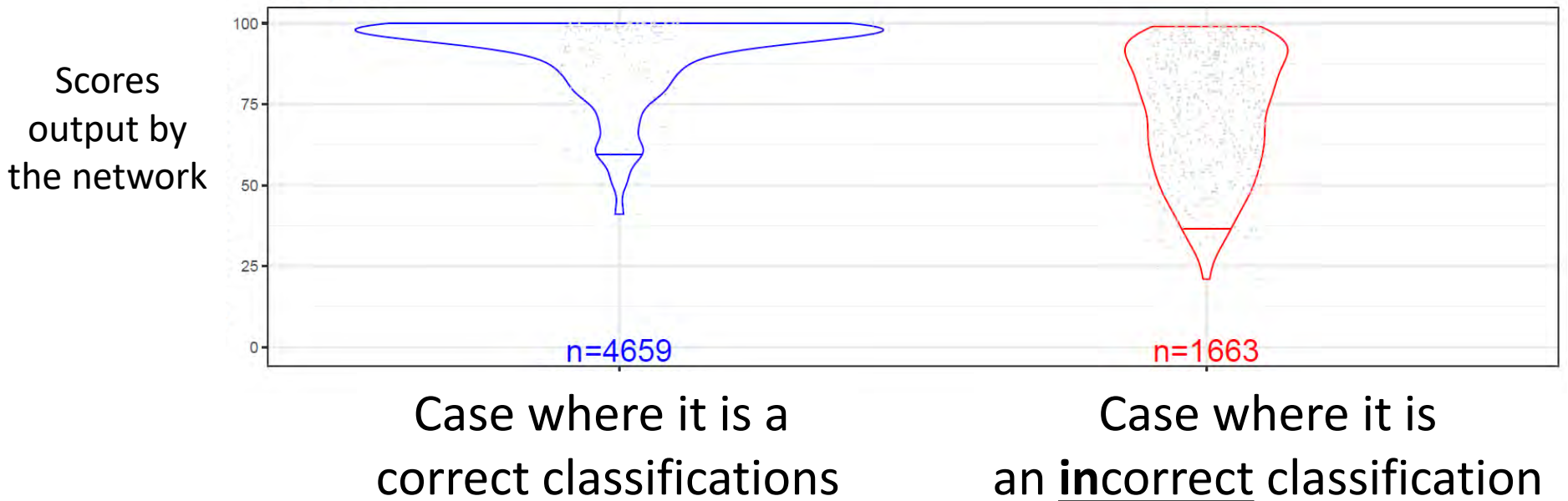
Result



The machine is 6% more accurate and 100 times faster

Can we really trust the results?

Output scores on a test set

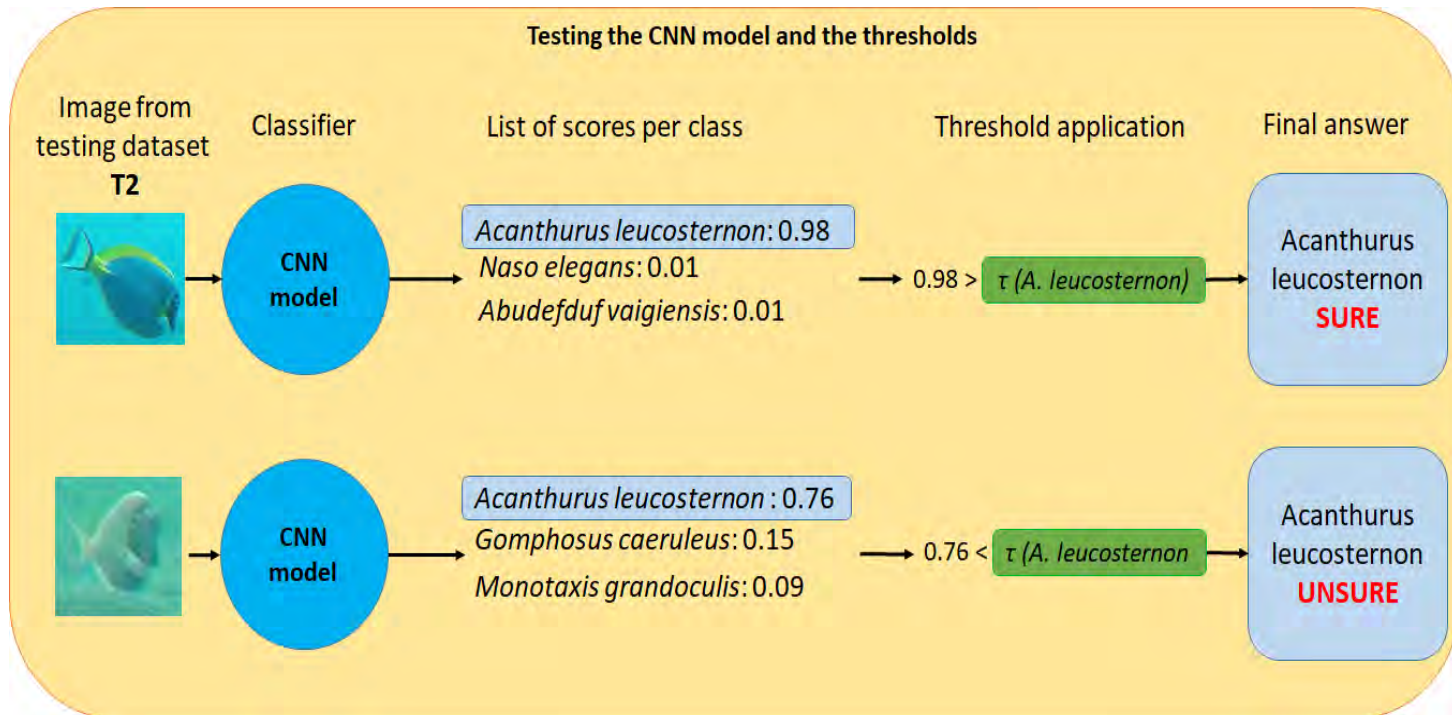


A possible solution

The post-processing:

The network can:

- predict a species,
- or can **refuse to predict (“unsure class”)**



Results

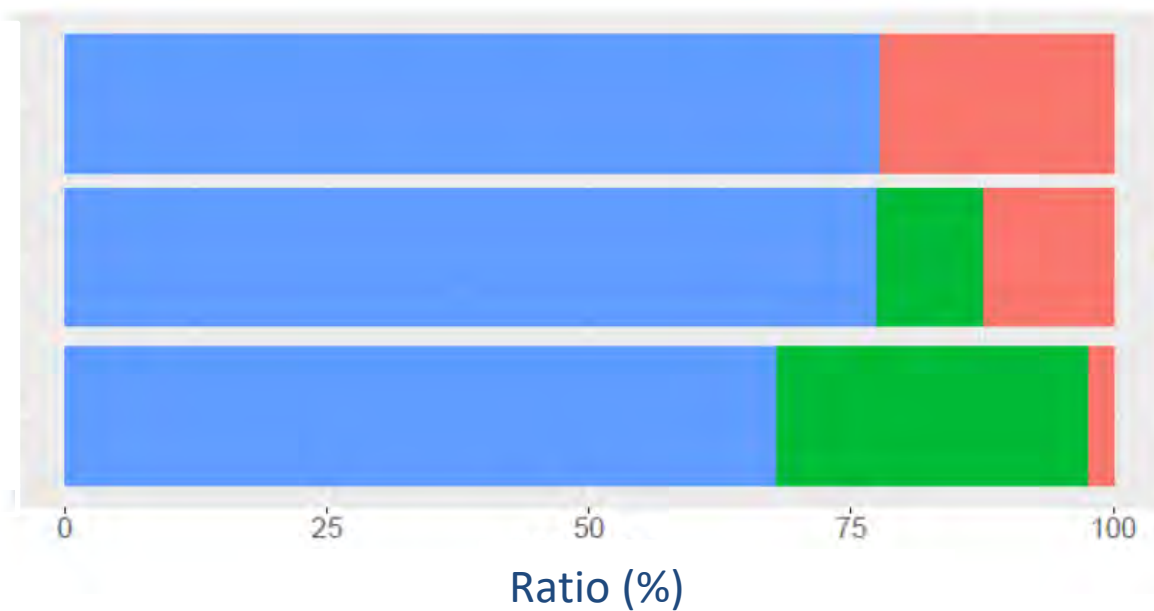
No post-processing

Objective n°1:

Keep correct
classification
maximum

Objective n°2:

Minimize the
incorrect
classification



Correct classification



Incorrect classification



« Unsure »

New project (in continuity)



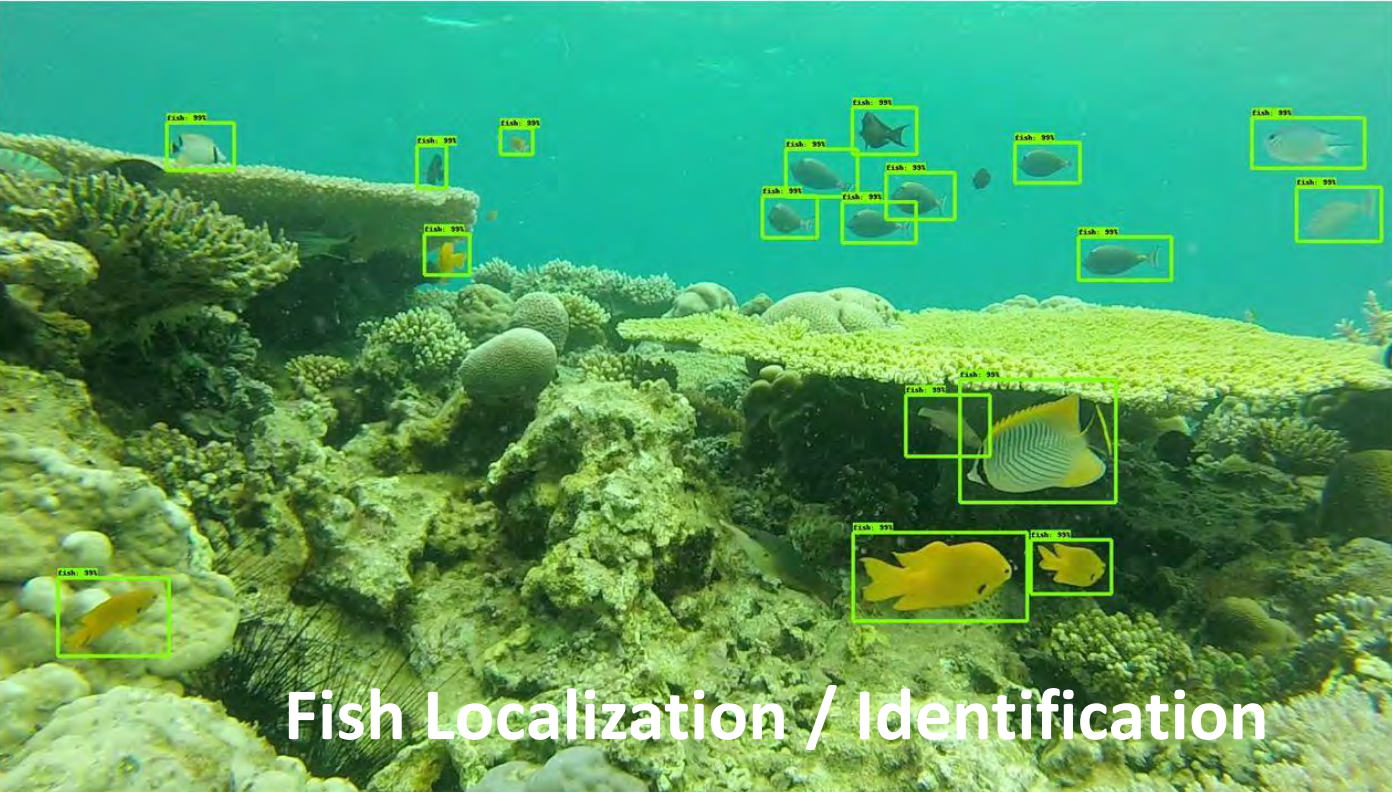
Stereo-System.....fixed on a.....robot

OBJECTIVE:

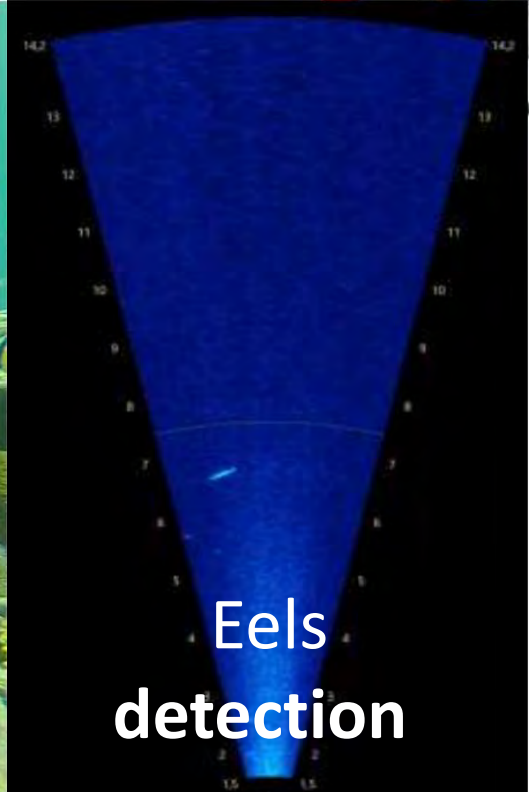
- Better detection,
- Size estimation,
- 3D motion estimation,
- Behavior estimation



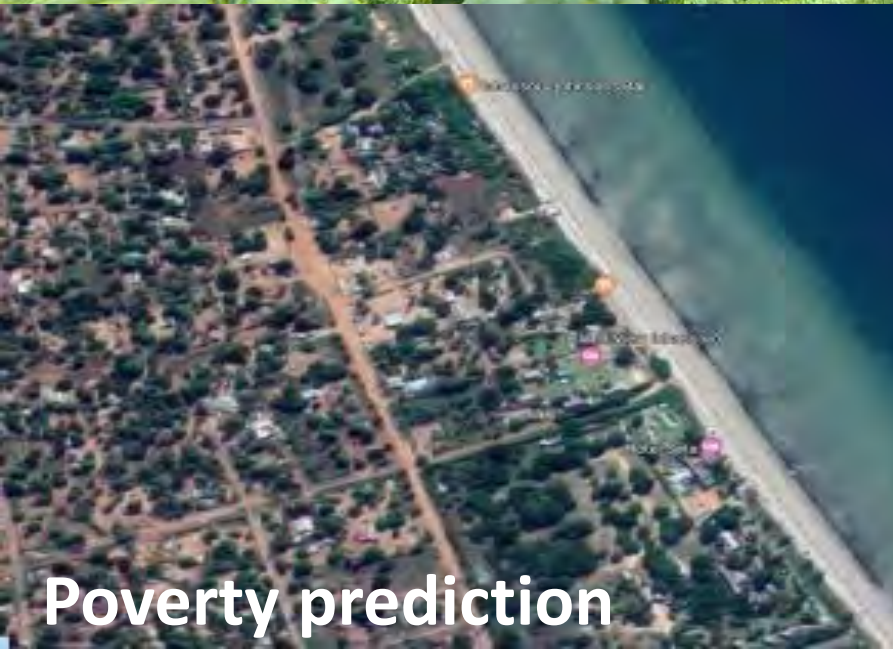
" Detection and counting of fish with a stereo-vision system",
research engineer: Sep. 2020 – August 2021 @LIRMM, Montpellier, ICAR.
Supervisors: Marc Chaumont, Gérard Subsol, ...



Fish Localization / Identification



Eels detection

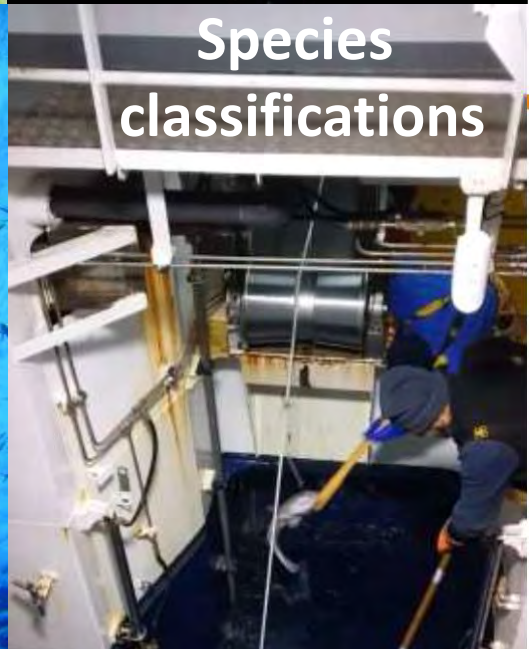


Poverty prediction



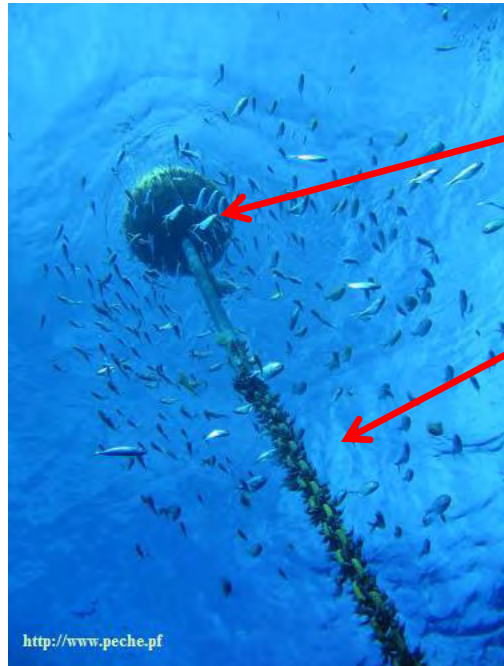
Shark localization
under DCP
with multiple
cameras

<http://www.peche.pf>



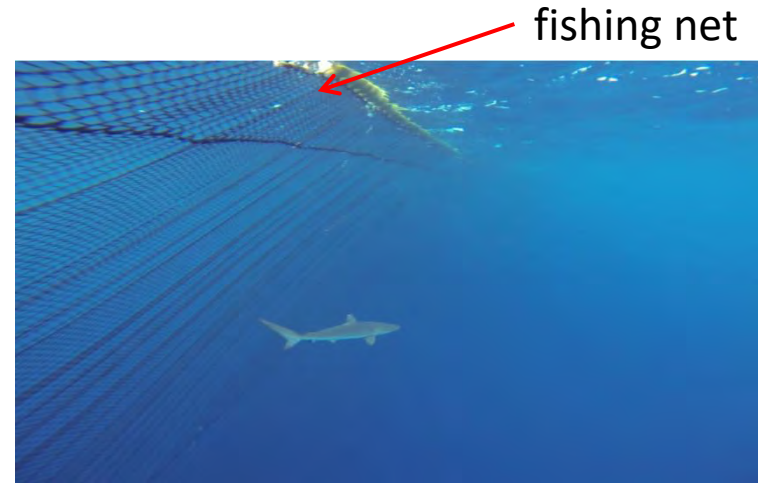
Species classifications

Shark localization // multi-view



Fish
Aggregating
Device
(FAP)

Chain



Shark



Tunas

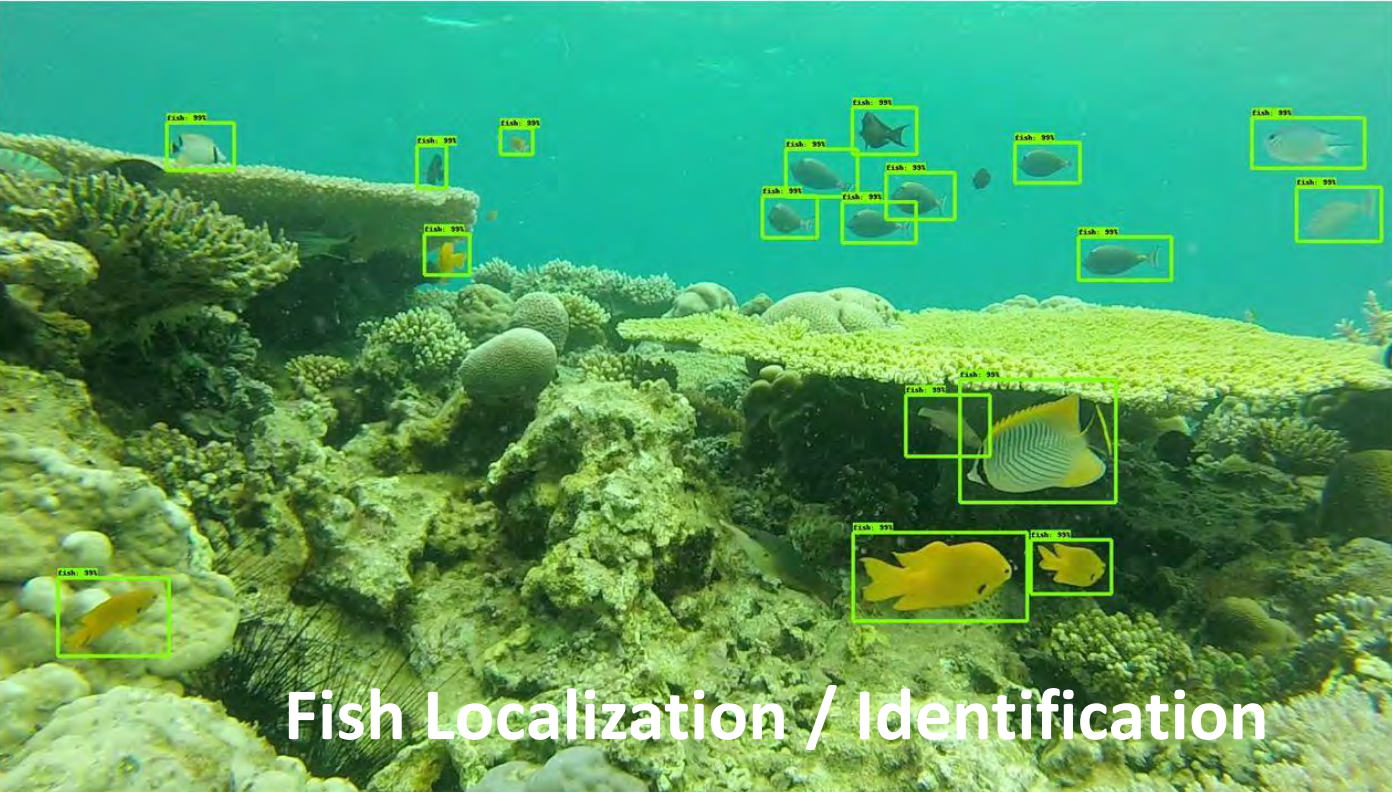


Container with
the cameras

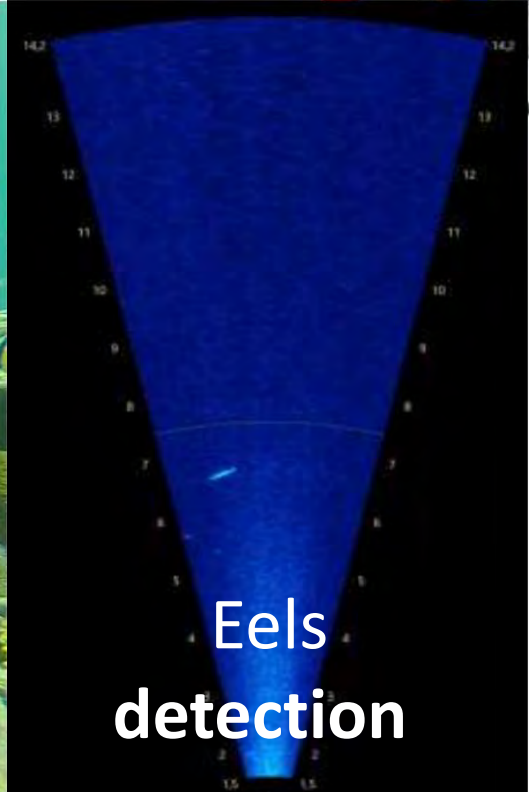
Prototype :
(this is just an
« artistic » vision...)

" Detection and counting of sharks from a sequence of multi-view underwater images by Deep-Learning ",
Post-Doc: Mehdi Yedroudj : Dec. 2019 – August 2020 @LIRMM, Montpellier, ICAR.

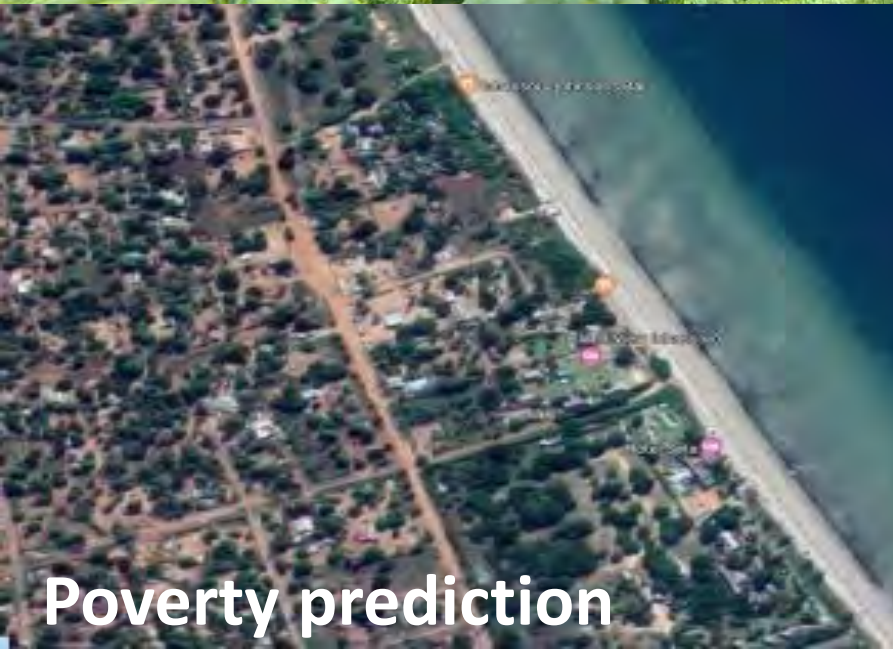
Supervisors: Marc Chaumont, Gérard Subsol, Vincent Creuze, Laurent Dagorn, Manuela Capello



Fish Localization / Identification



Eels
detection



Poverty prediction



Shark
localization

under DCP
with multiple
cameras

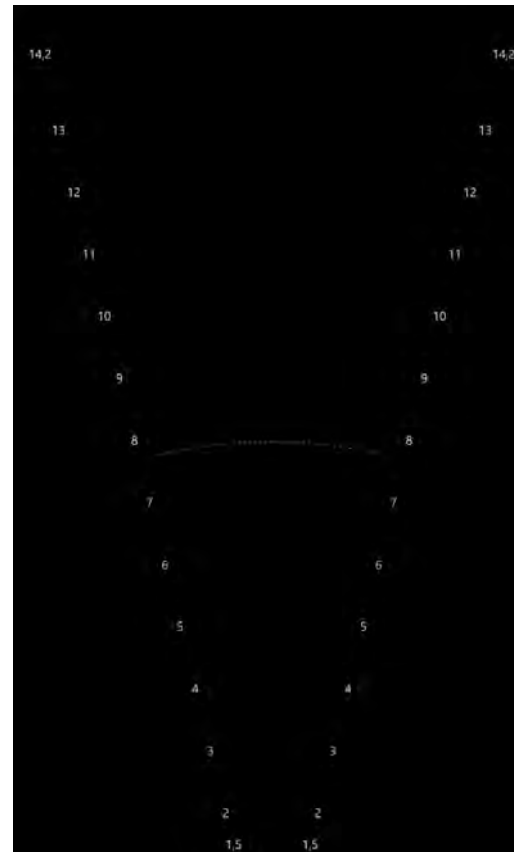
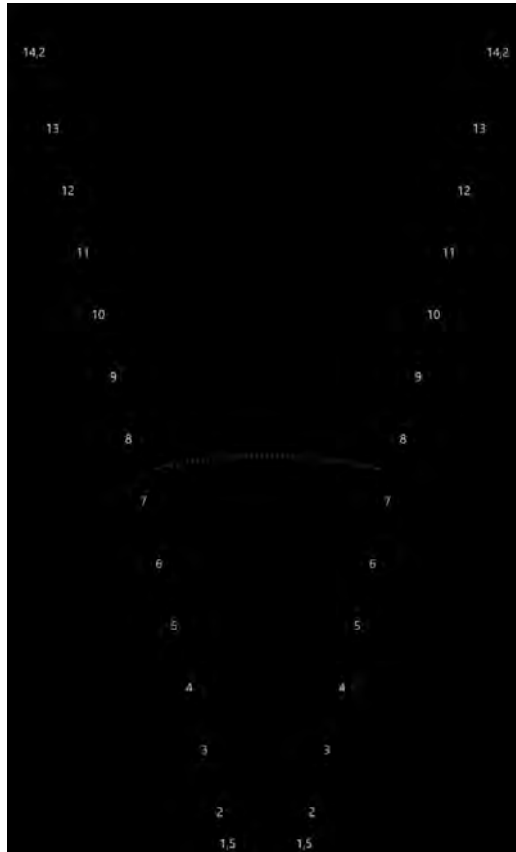
<http://www.peche.pf>



Species
classifications

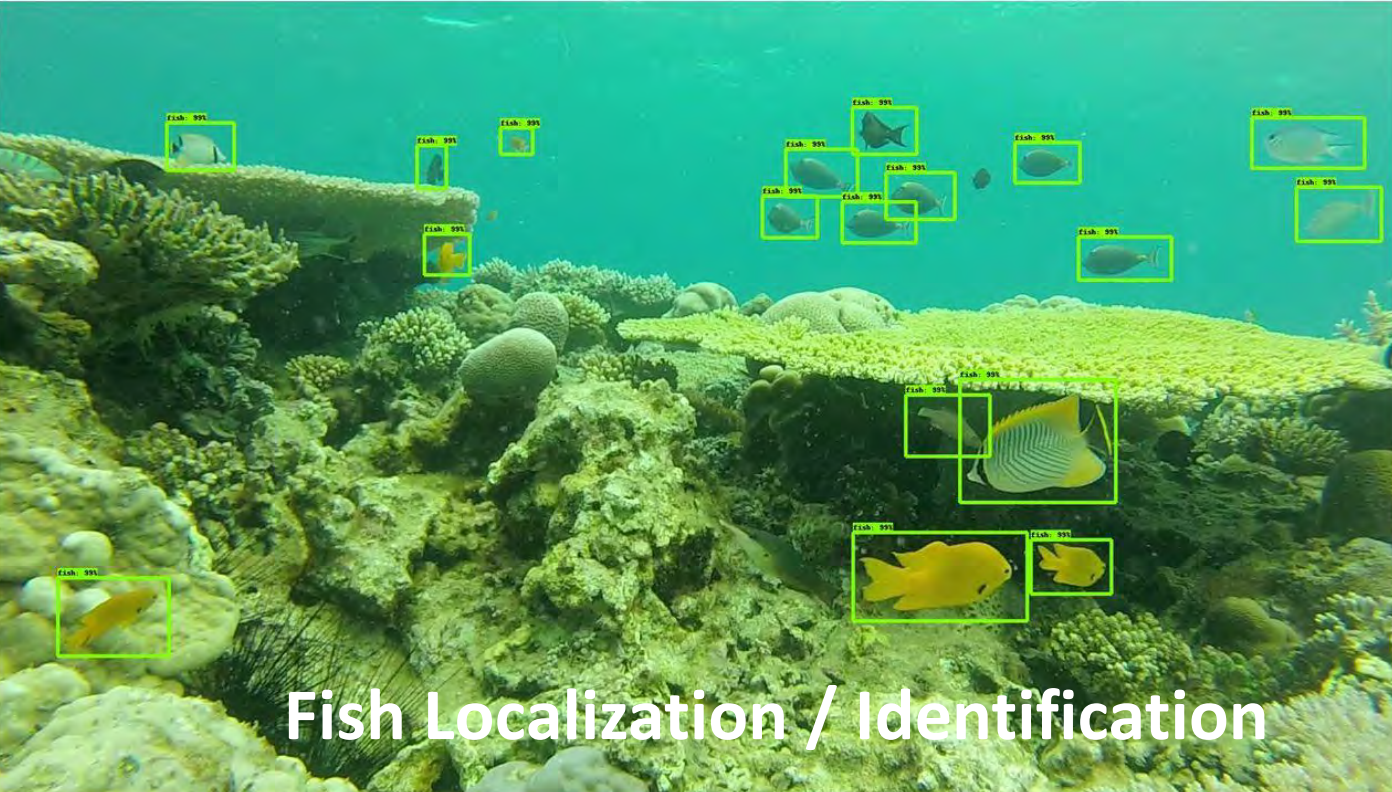
Eels counting

European management plan for the restoration of the eel stock

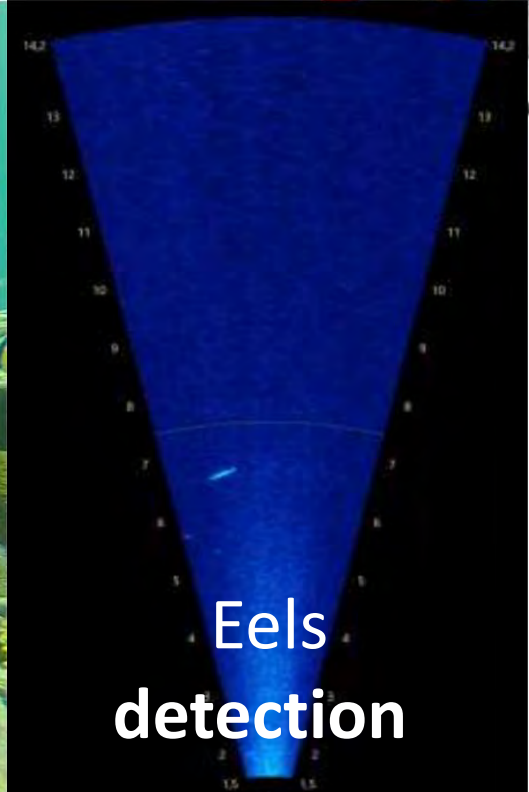


**" [Identification and counting of eels from multibeam sonar videos by Deep-Learning](#) ",
Master 2 internship **2020** @LIRMM, Montpellier, ICAR.**

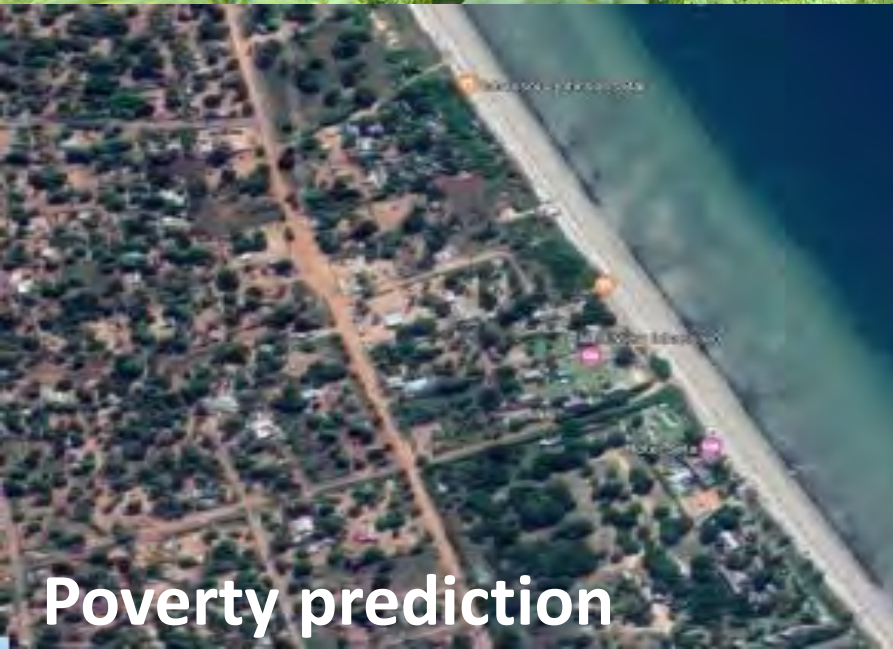
Supervisors: Gérard Subsol, Vincent Creuze, Mehdi Yedroudj, Marc Chaumont, Jason Peyre, Raphaël Lagarde



Fish Localization / Identification



Eels
detection

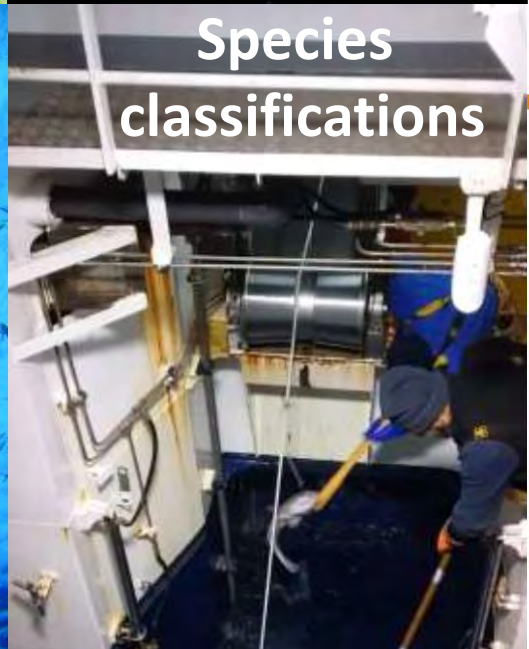


Poverty prediction



Shark
localization
under DCP
with multiple
cameras

<http://www.peche.pf>



Species
classifications

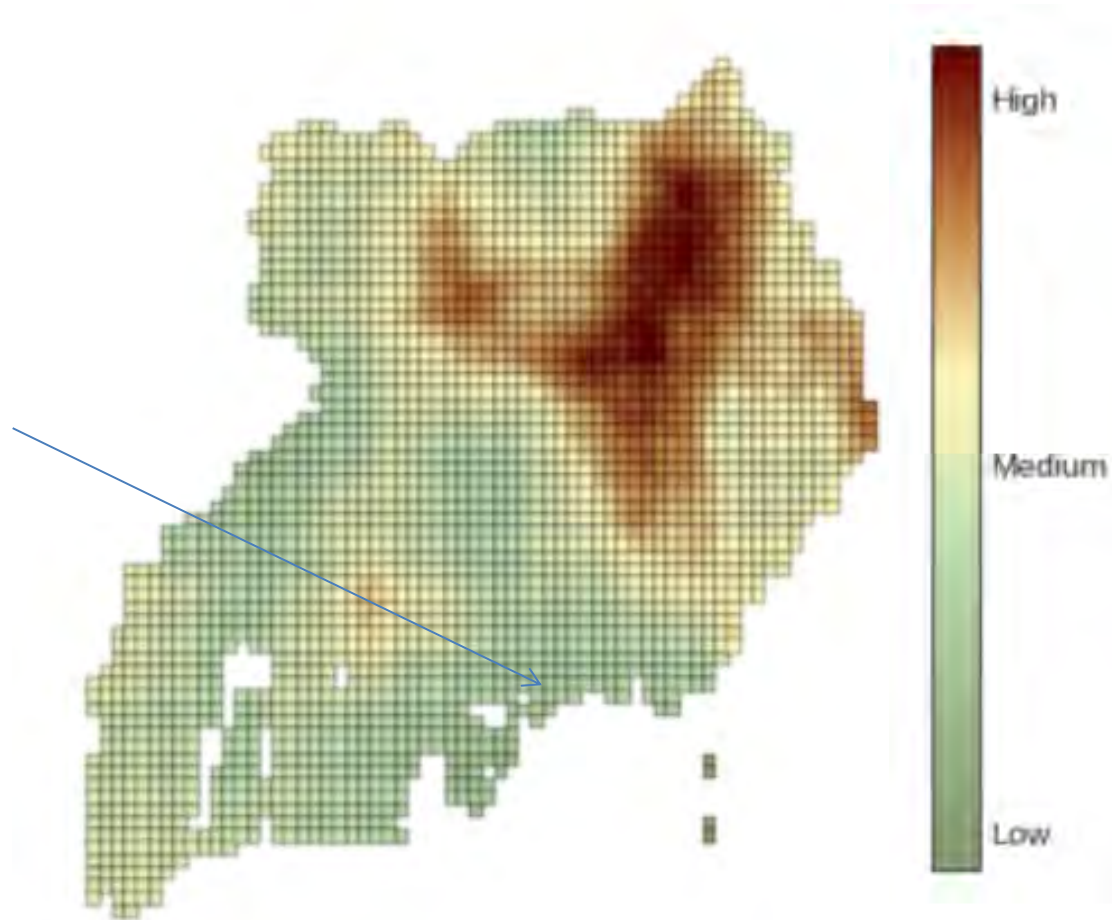
Prediction of poverty with a CNN ?

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 \approx 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 computer-science research?

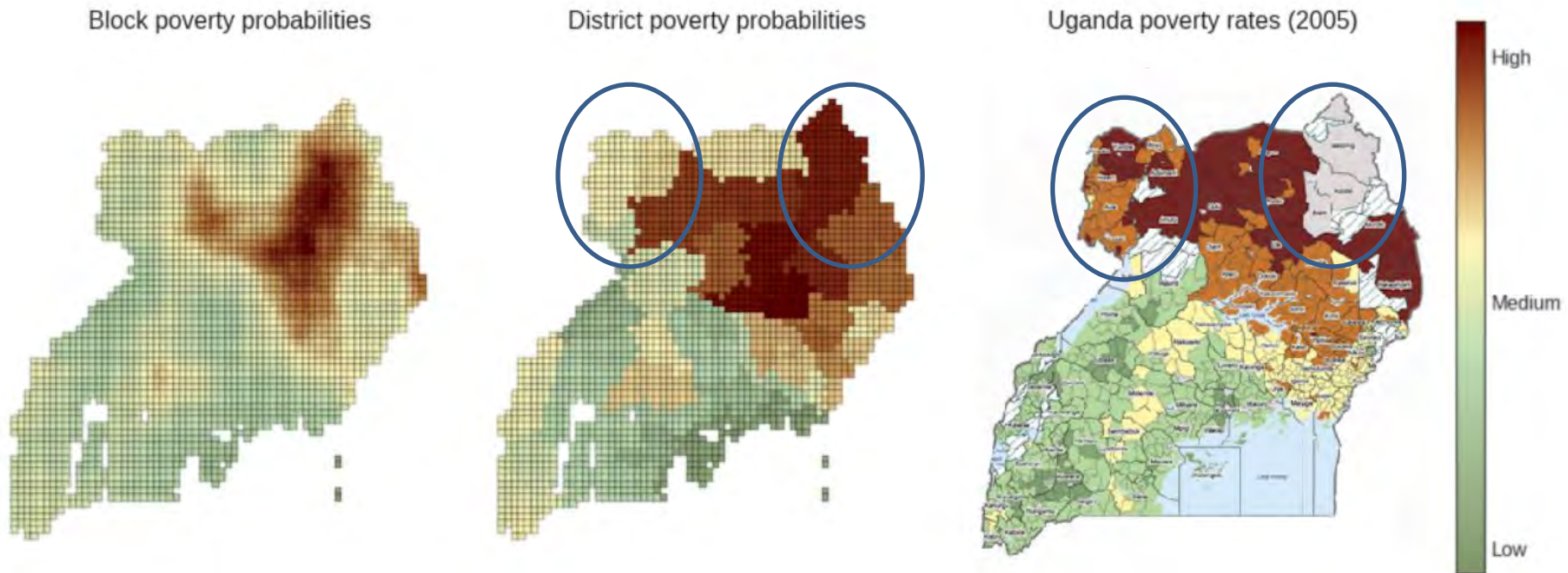
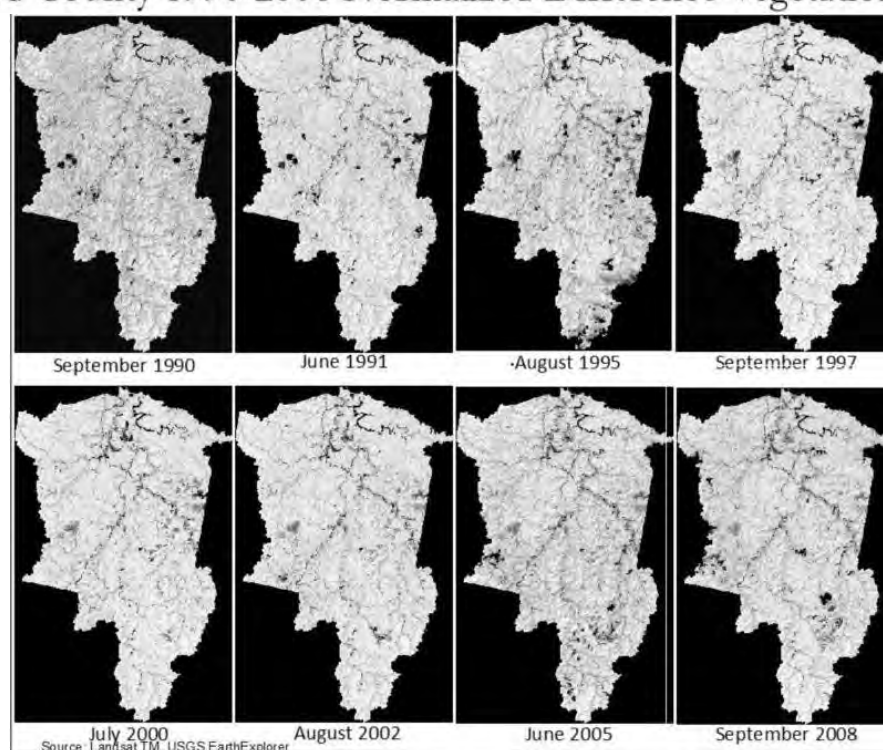


Figure 3: **Left:** Predicted poverty probabilities at a fine-grained 10km × 10km block level. **Middle:** Predicted poverty probabilities aggregated at the district-level. **Right:** 2005 survey results for comparison (World Resources Institute 2009).

Only correlated to 70% to the ground truth

Objective: work with images sequences

Floyd County 1990-2008 Normalized Difference Vegetation Index



Deep learning

→ Poverty (0% ... 100%)

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

<https://ericvenson.com/monitoring-mining-impact/>

Post-doc (funded by Belmont/CESAB) 2020

Thesis (ANR) in September 2020

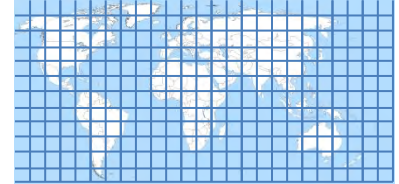
" [Poverty prediction by Deep Learning from a satellite image sequence](#) ",
Master 2 internship **2020** @LIRMM, Montpellier, ICAR.
Supervisors: Marc Chaumont, Gérard Subsol, Laure Berti-Équille, Dino Ienco

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 2020



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

Master internship 2021 ?



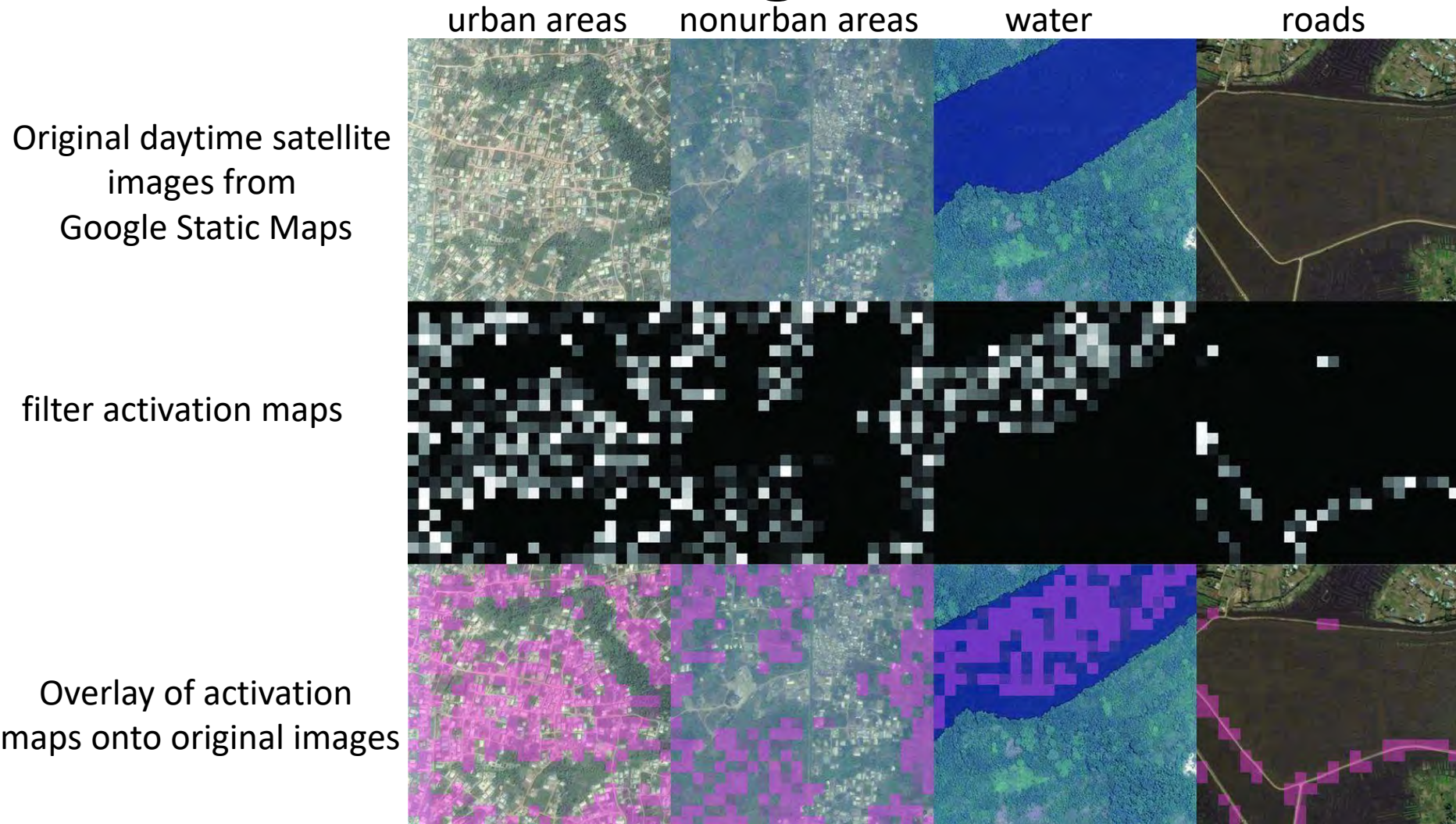
SUBJECTS discussed in the past ... :

- **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...



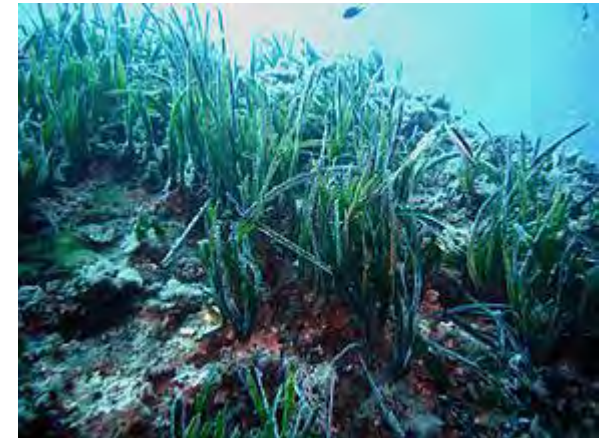
TO BE CONTINUED ...

Parts of the image that “react”



Comptage du nombre de bateaux

- La baie de Paulilles (aire protégée)



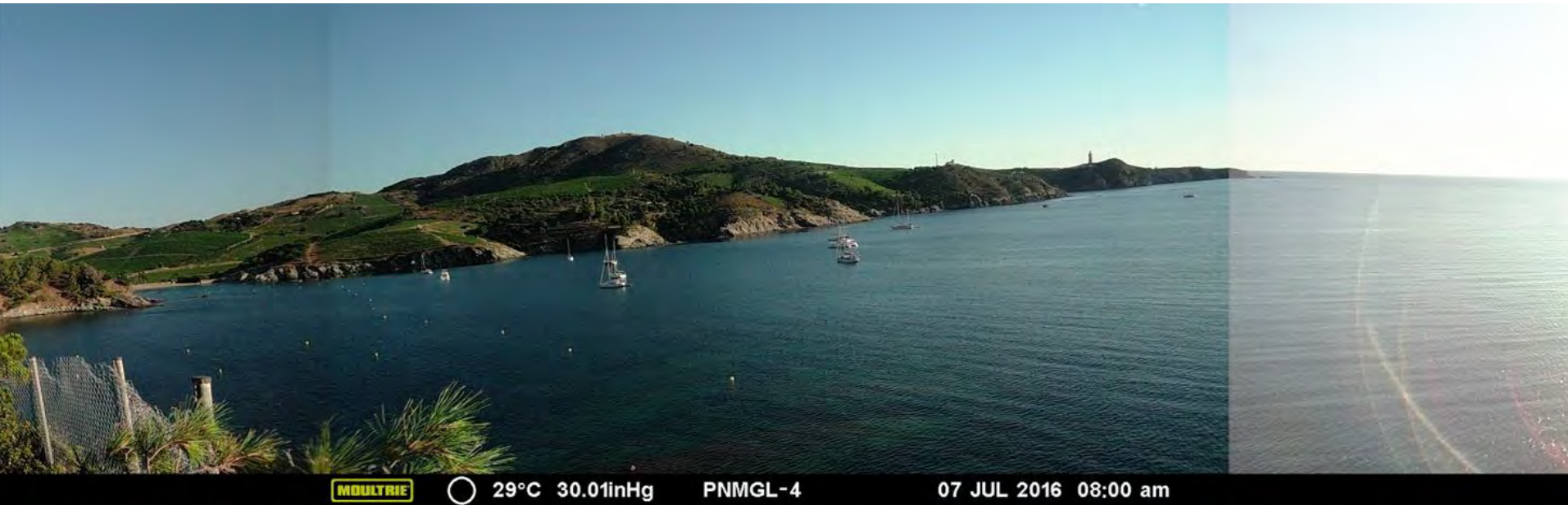
Posidonies



Coralligène = éco-syst. avec algues calcaire

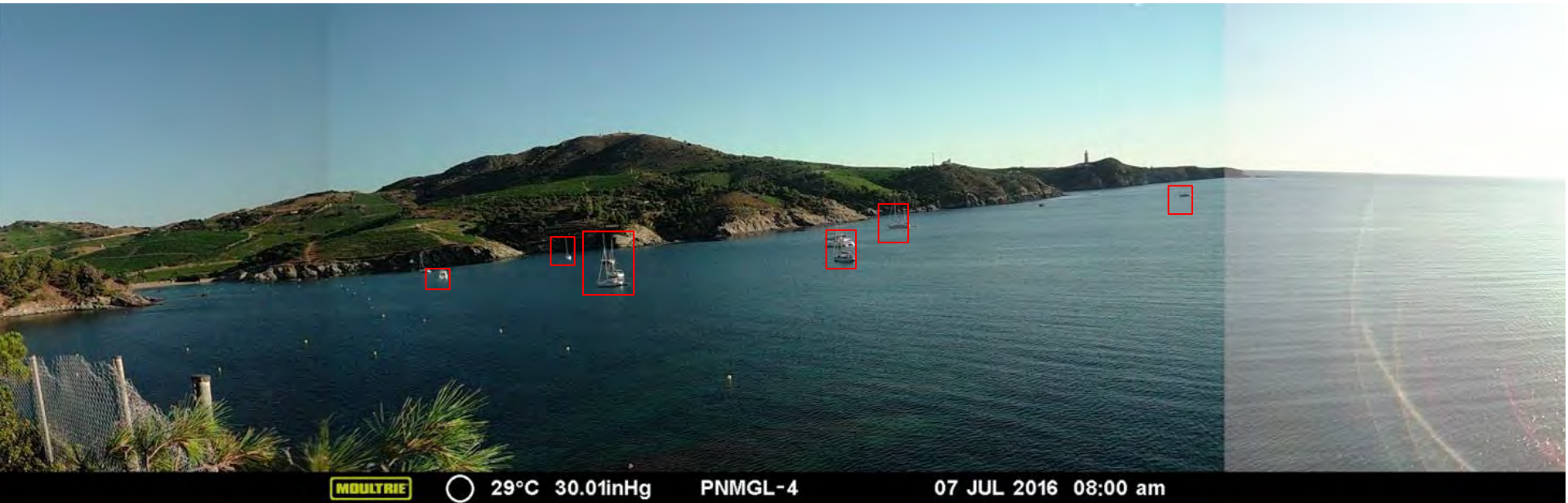
- Question du parc marin :
Corrélation nb bateau / dégradation fond marin ?

Dispositif de surveillance



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

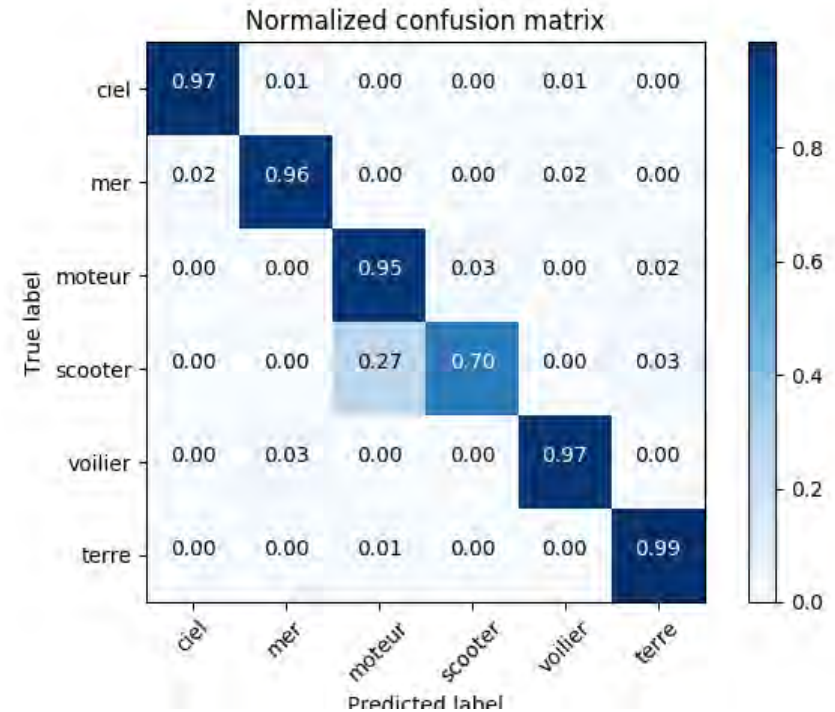
Ce que l'on voudrait obtenir



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).

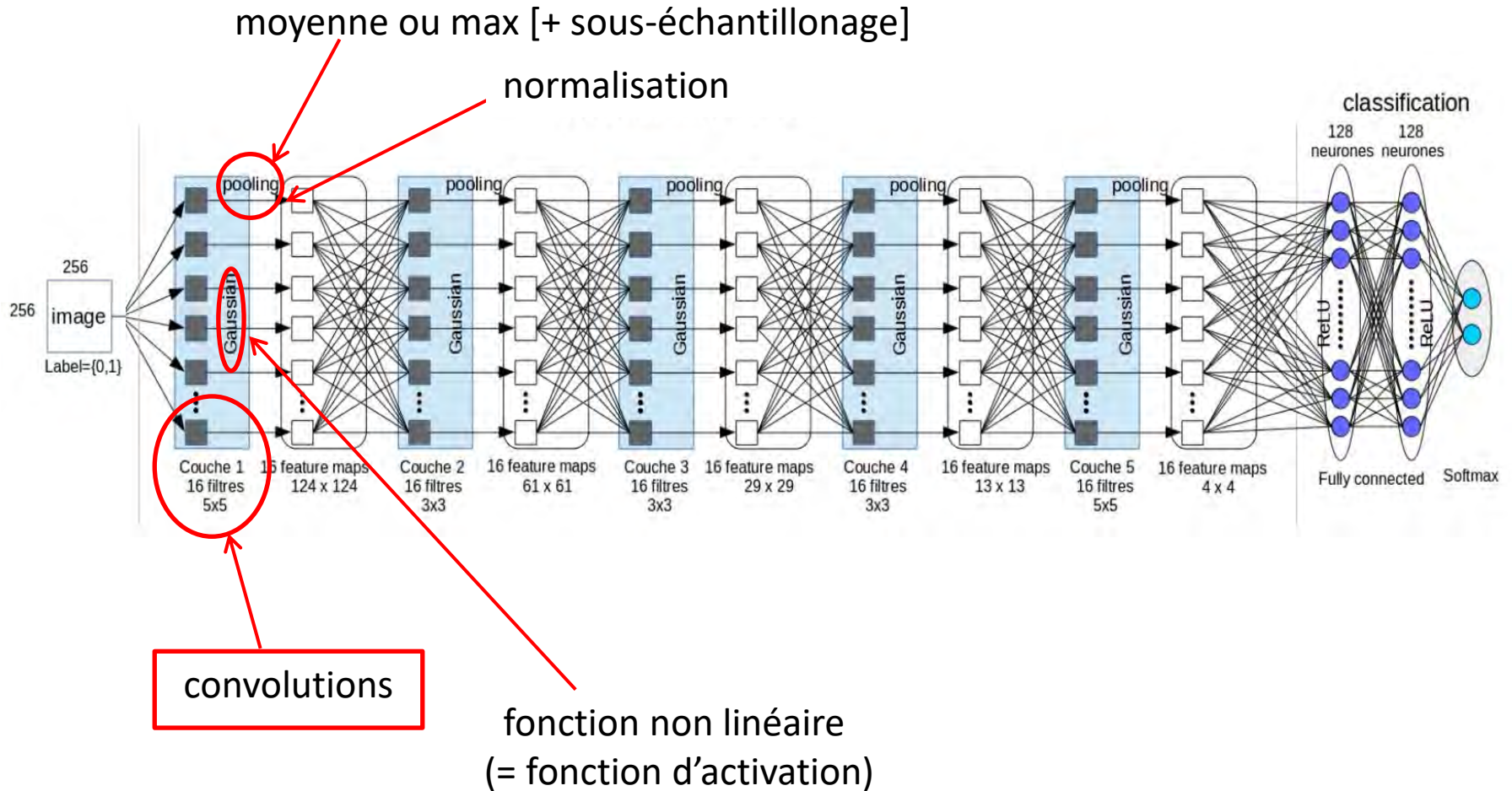
Inception v3 pré-entraîné sur 2 millions d'images sur ImageNet (transfer learning)

That's all folks!

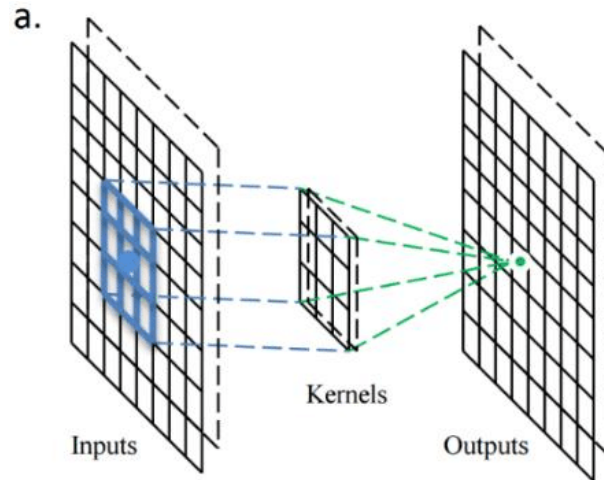


Le réseau CNN

Convolutional Neural Network



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

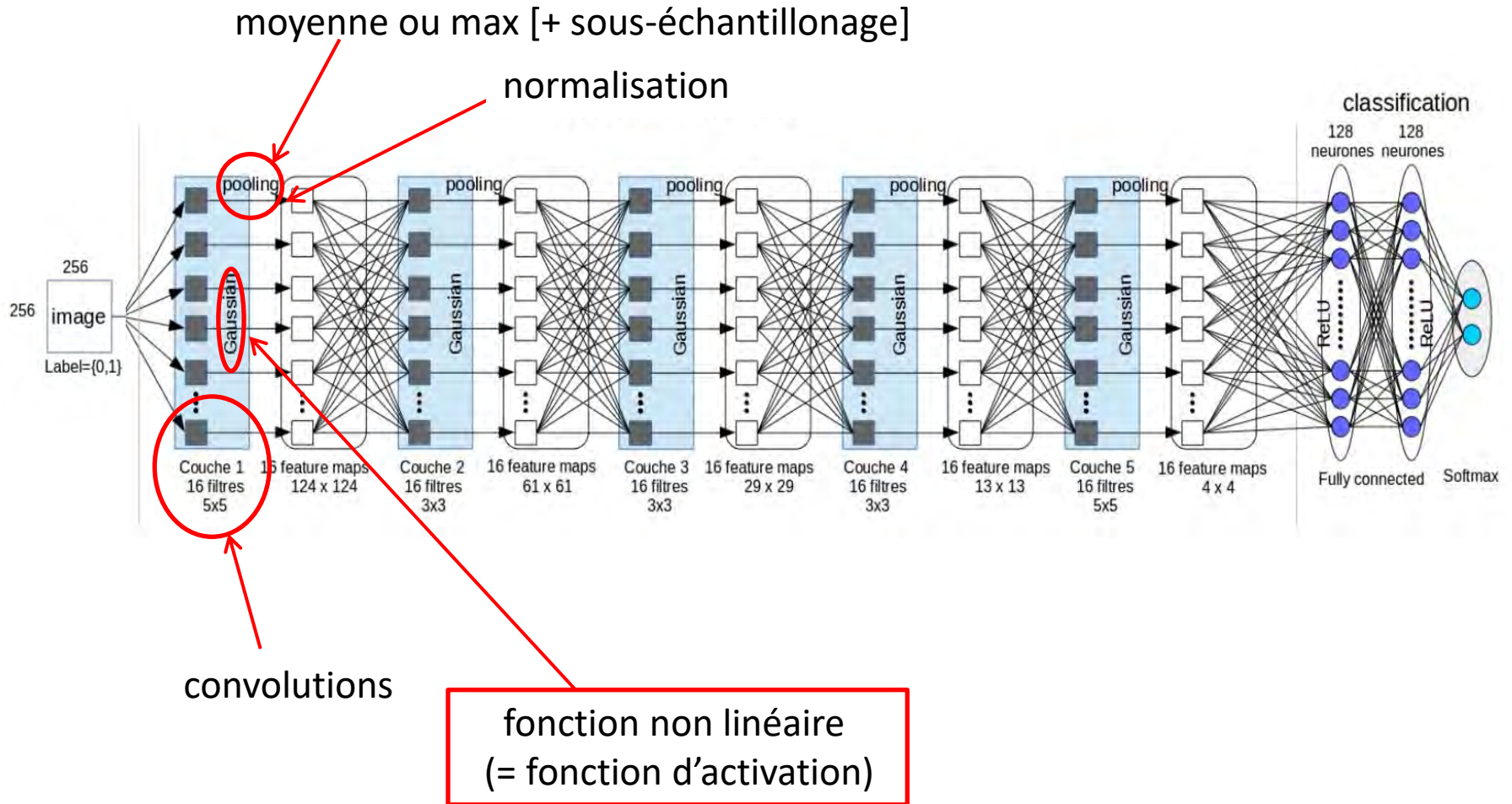
« kernel » (noyau)

?	?	?
?	114,4	?
?	?	?

résultat (output)

Le réseau CNN

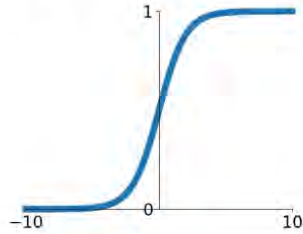
Convolutional Neural Network



Activation Functions

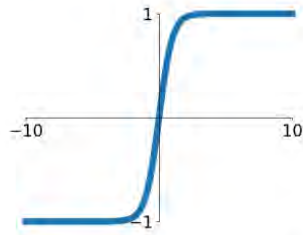
Sigmoid

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



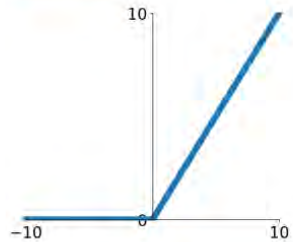
tanh

$$\tanh(x)$$



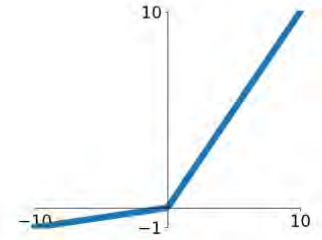
ReLU

$$\max(0, x)$$



Leaky ReLU

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

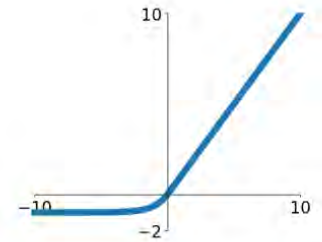


Maxout

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

ELU

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

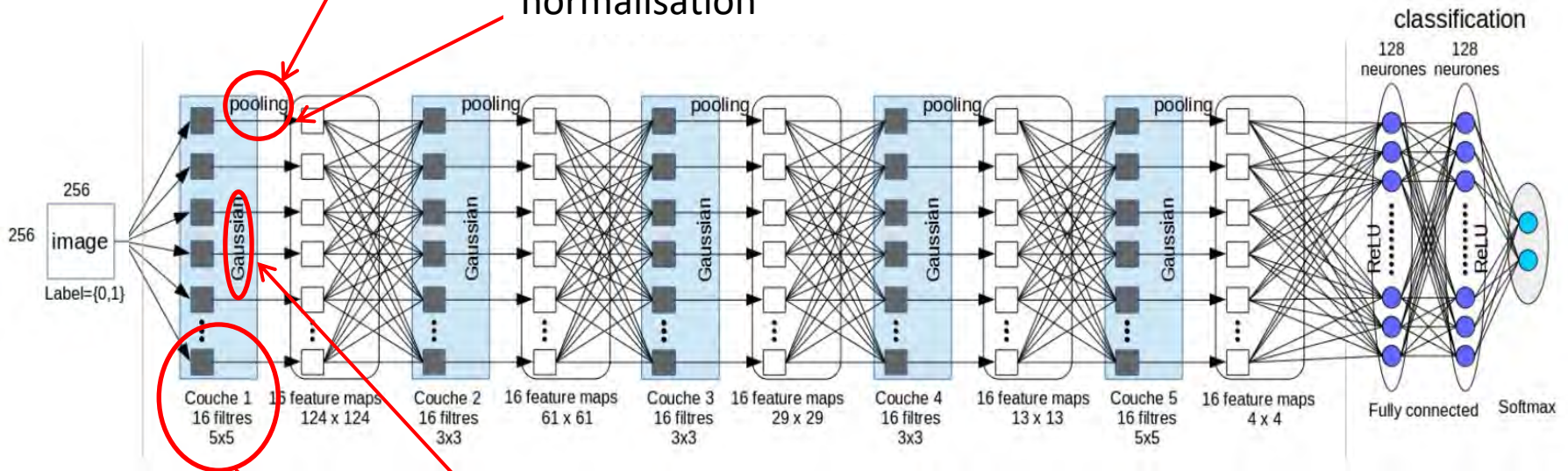


Le réseau CNN

Convolutional Neural Network

moyenne ou max [+ sous-échantillonnage]

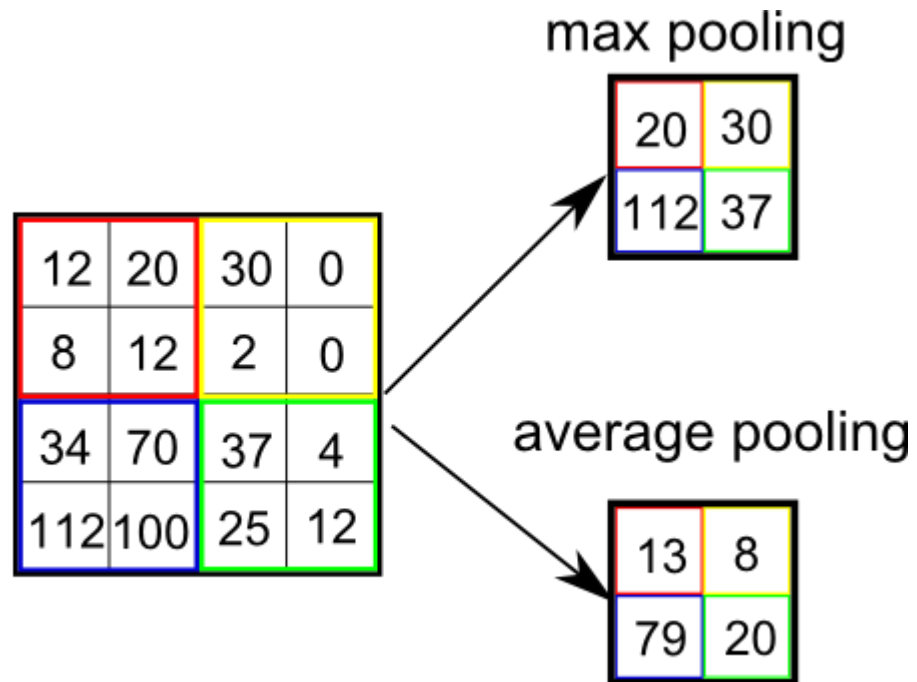
normalisation



convolutions

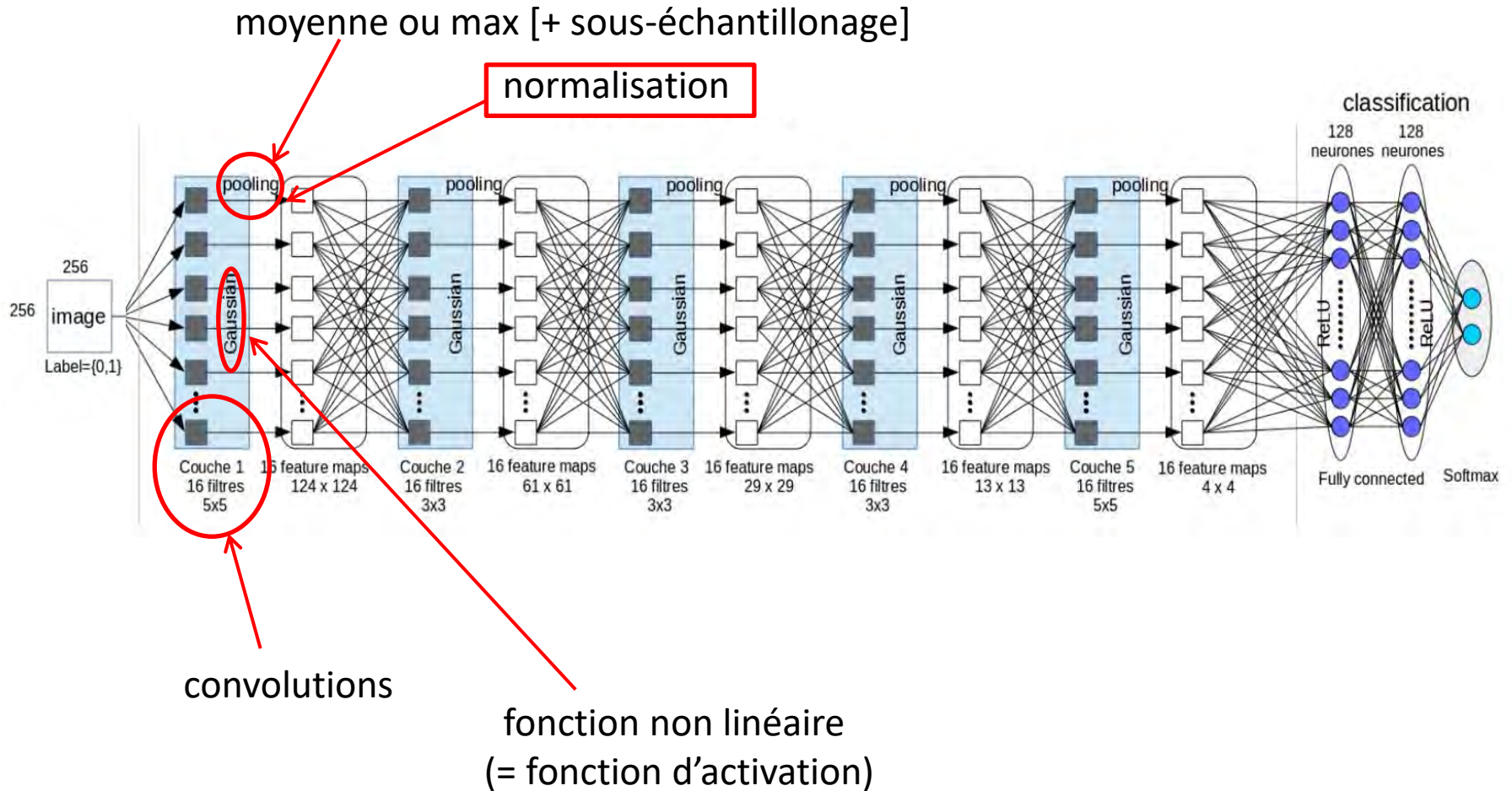
fonction non linéaire
(= fonction d'activation)

Le pooling

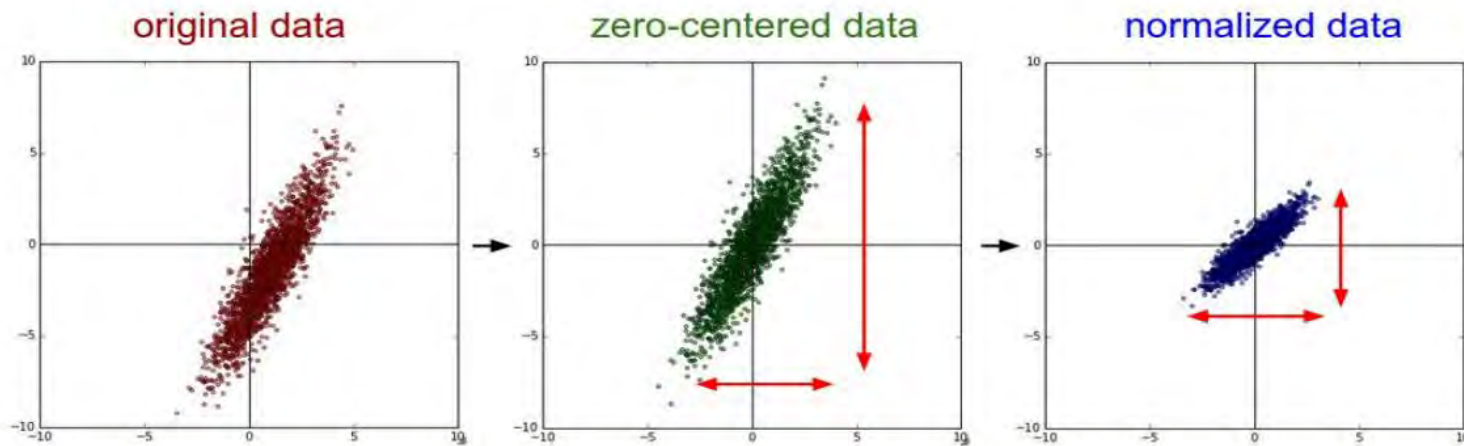


Le réseau CNN

Convolutional Neural Network



Exemple : Batch Normalisation

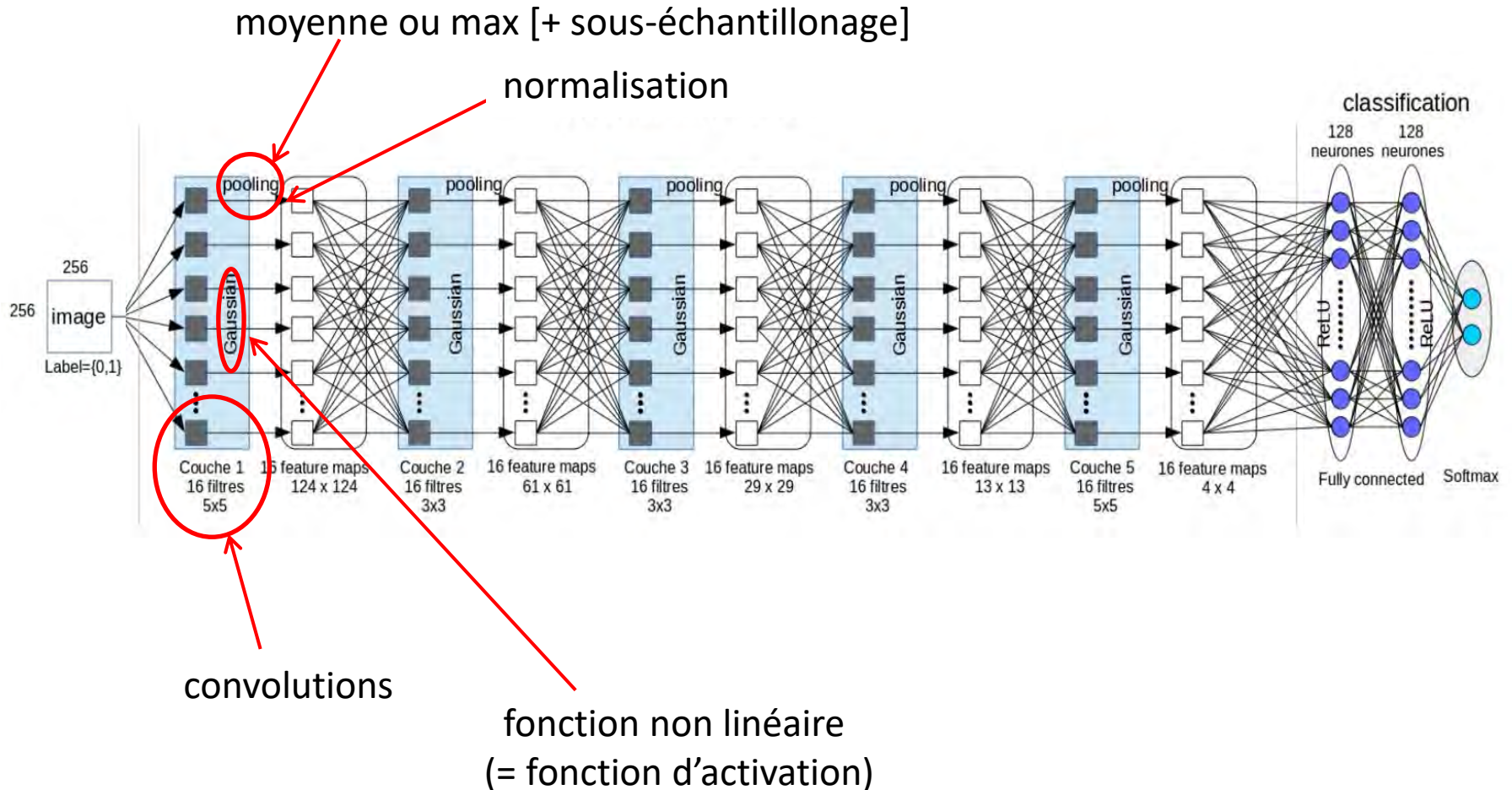


Batch Normalization

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

Le réseau CNN

Convolutional Neural Network

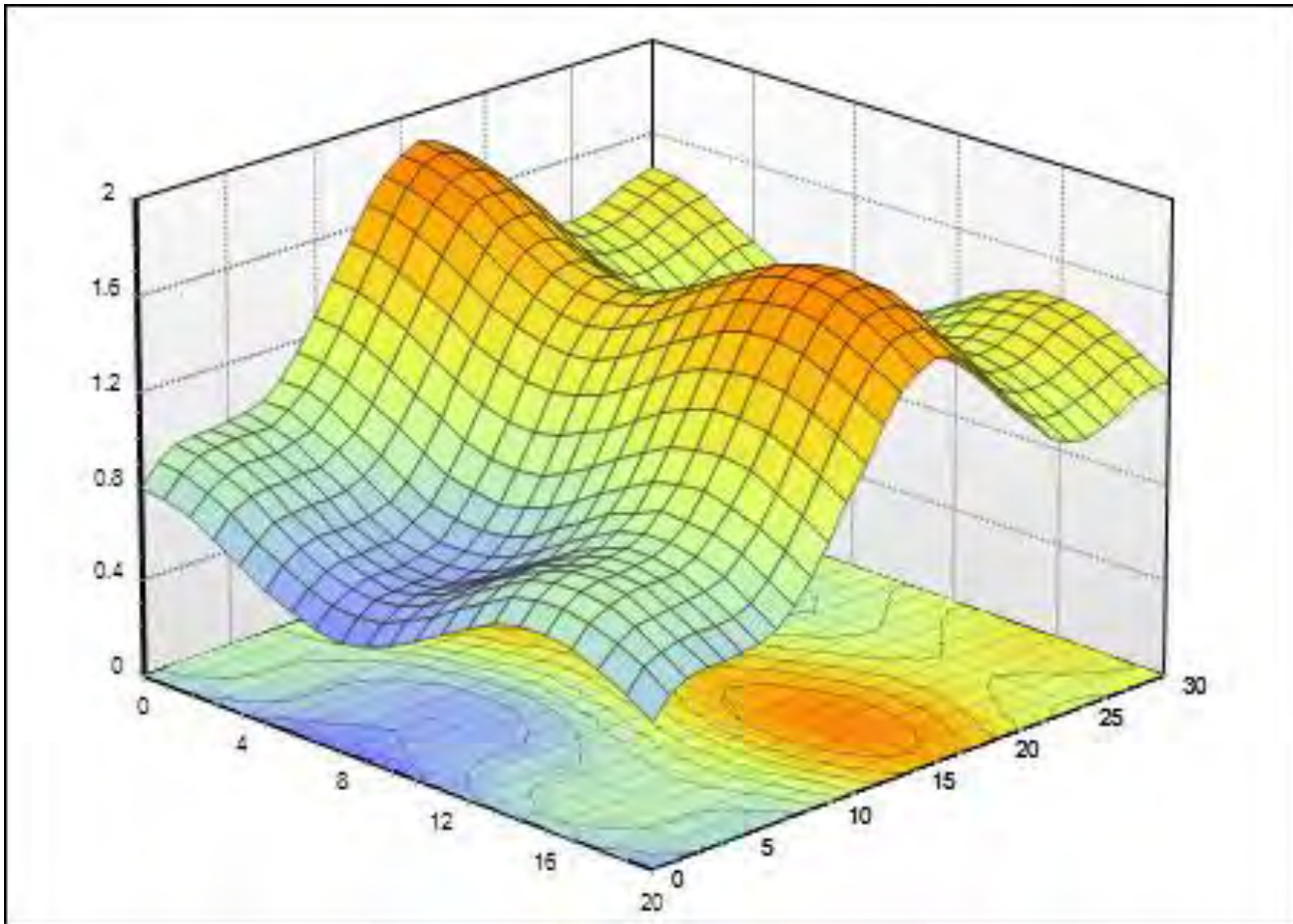


Formellement
(et grossièrement..)

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

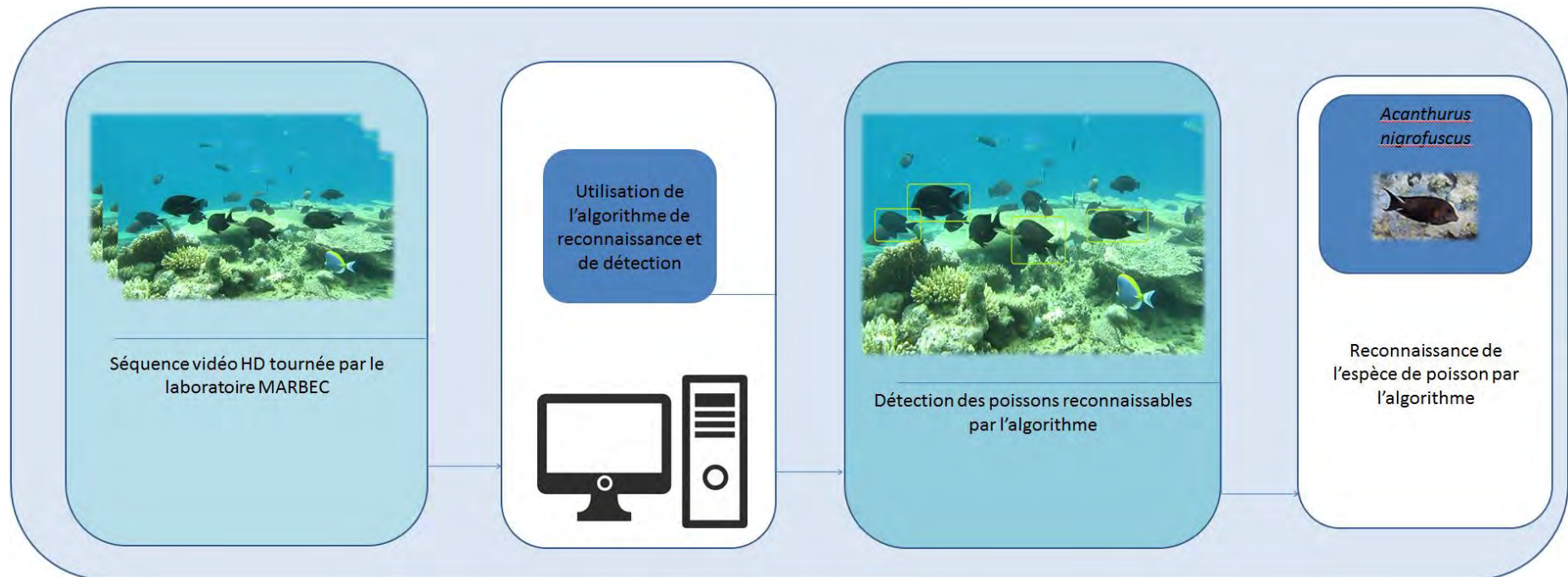
L'apprentissage (expliqué avec les mains..)

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, ...

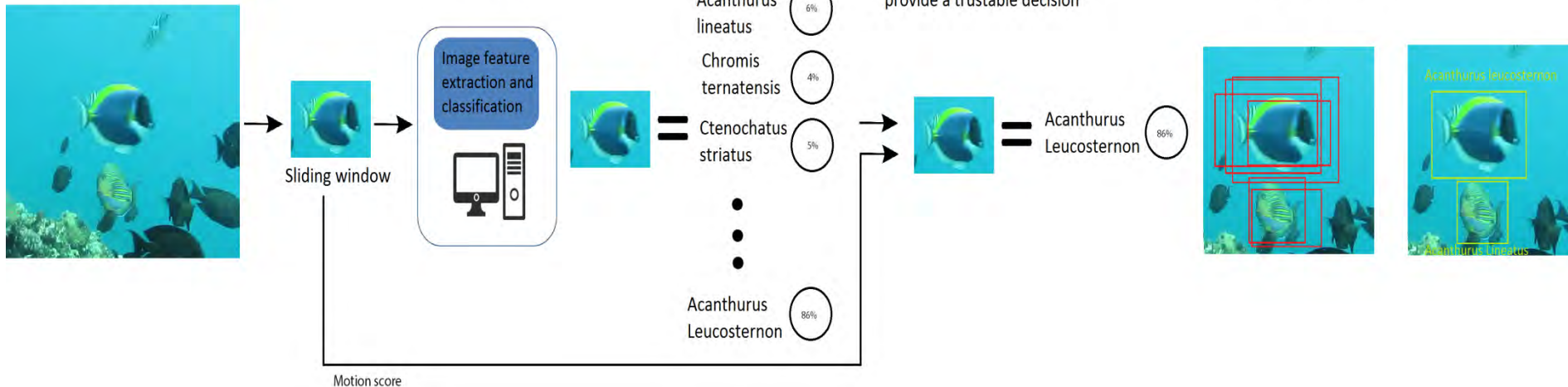
L'architecture globale

At a given resolution, pass a multi-resolution sliding window through the frame

The detection and recognition method computes probabilities for the window to belong to different classes (fish species or background)

According to some thresholds, we decide if the probabilities and motion scores are high enough to provide a trustable decision

We finally fuse bounding boxes to provide a final decision about localization and identification

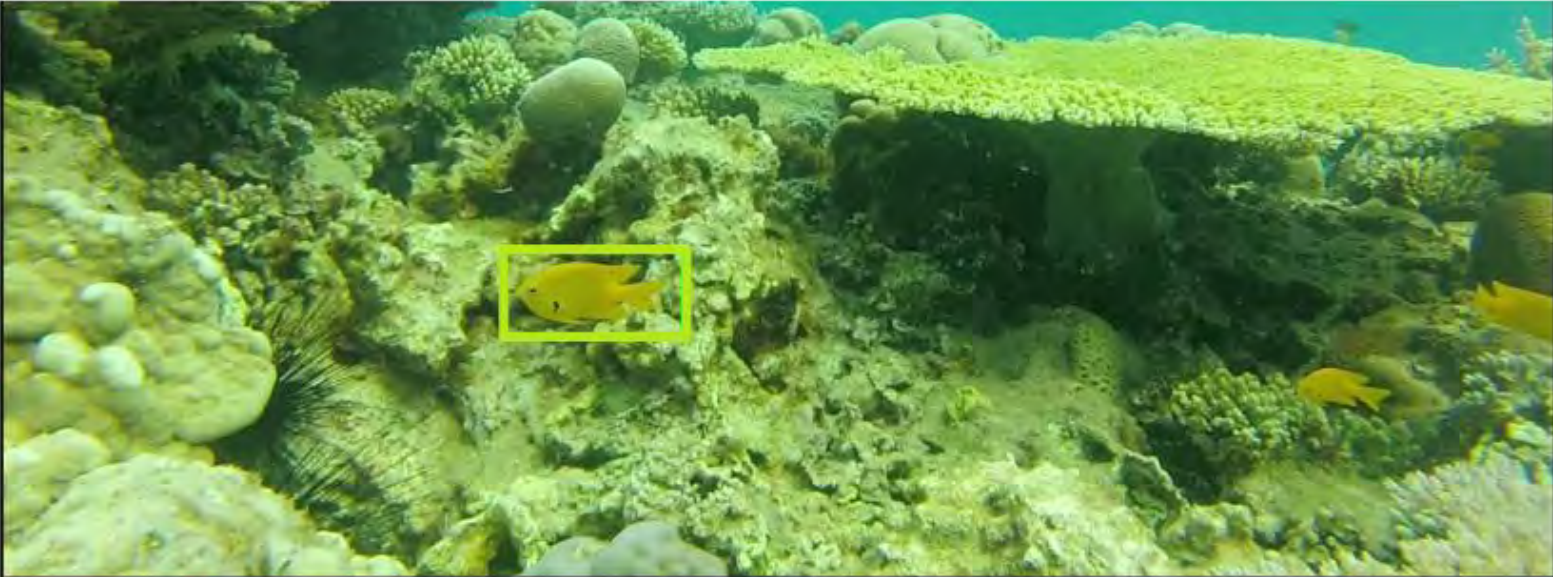


GoogLeNet avec 27 couches et un soft-max.


Sebastien Villon, Marc Chaumont, Gerard Subsol, Sebastien Villegier, 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 ».



Select a file ...

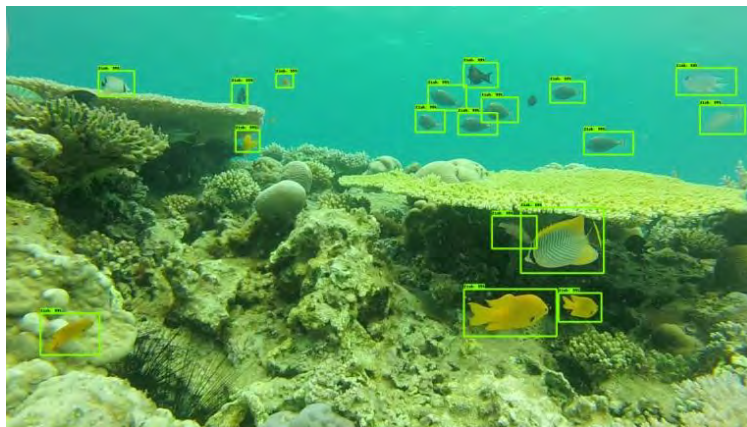


Who is?

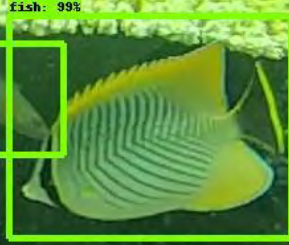
information	Classification took 0.0319459438324 seconds
Score	label
99.9515%	Pomacentrus sulfureus
0.0204%	Acanthurus lineatus
0.0183%	Pseudanthias squamipinnis female
0.0067%	Chlorurus gibbus female
0.0025%	Chromis ternatensis
Time	1.55859206107 Second

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)

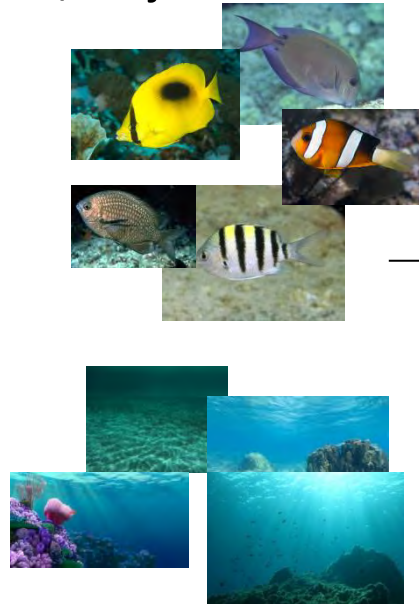


Classification automatique à partir d'images

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



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

Humain vs Machine

tache de classification d'images

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

Images testées

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

Taux de bonne classification

- Temps identification :
 - 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

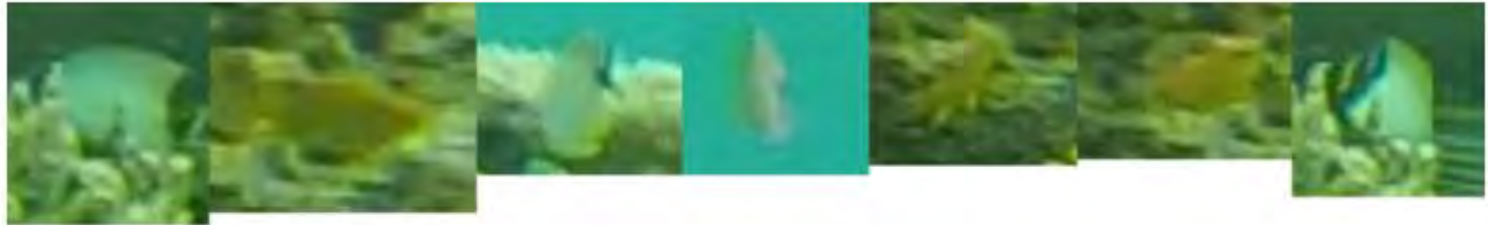


Figure 5: Sample of pictures recognized by the network and not recognized by experts.



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

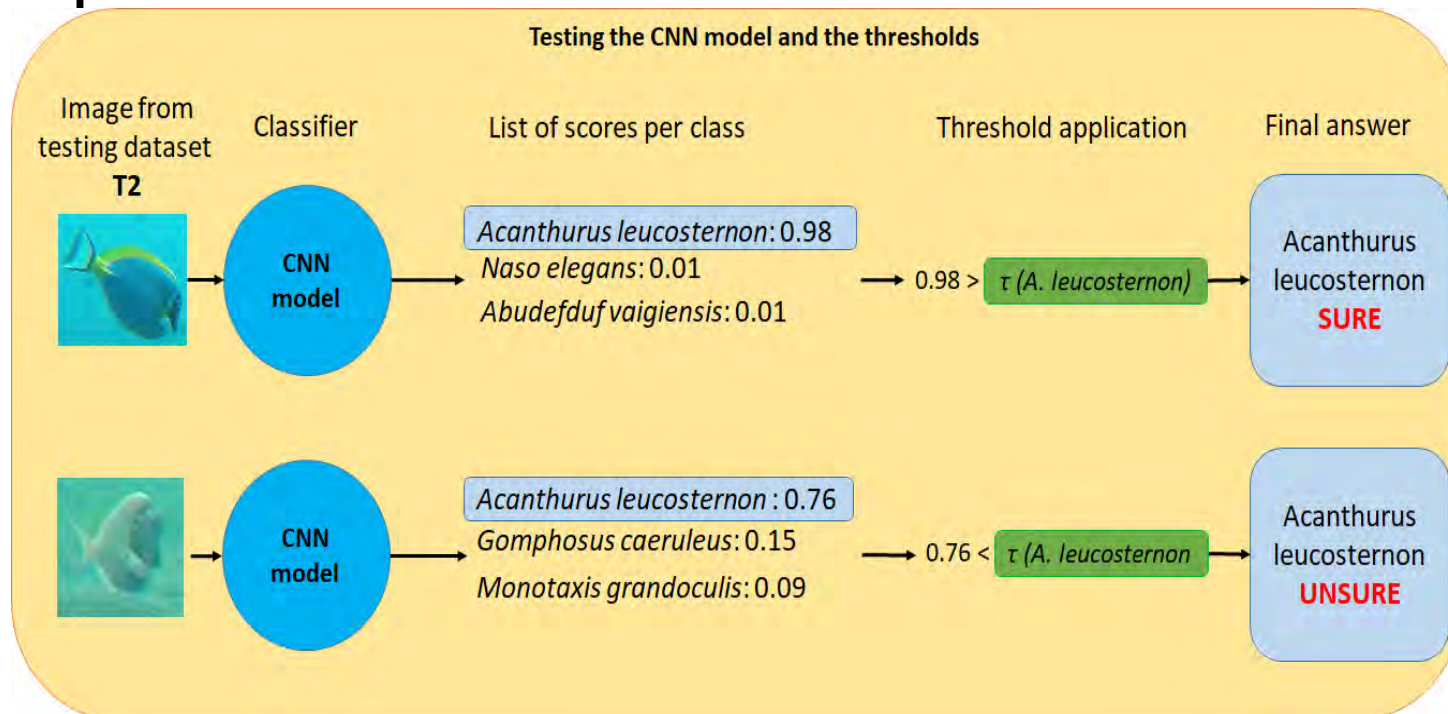


Figure 7: Sample of pictures misclassified by both humans and network.

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

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

- Seuil 2 : borne l'erreur de classif

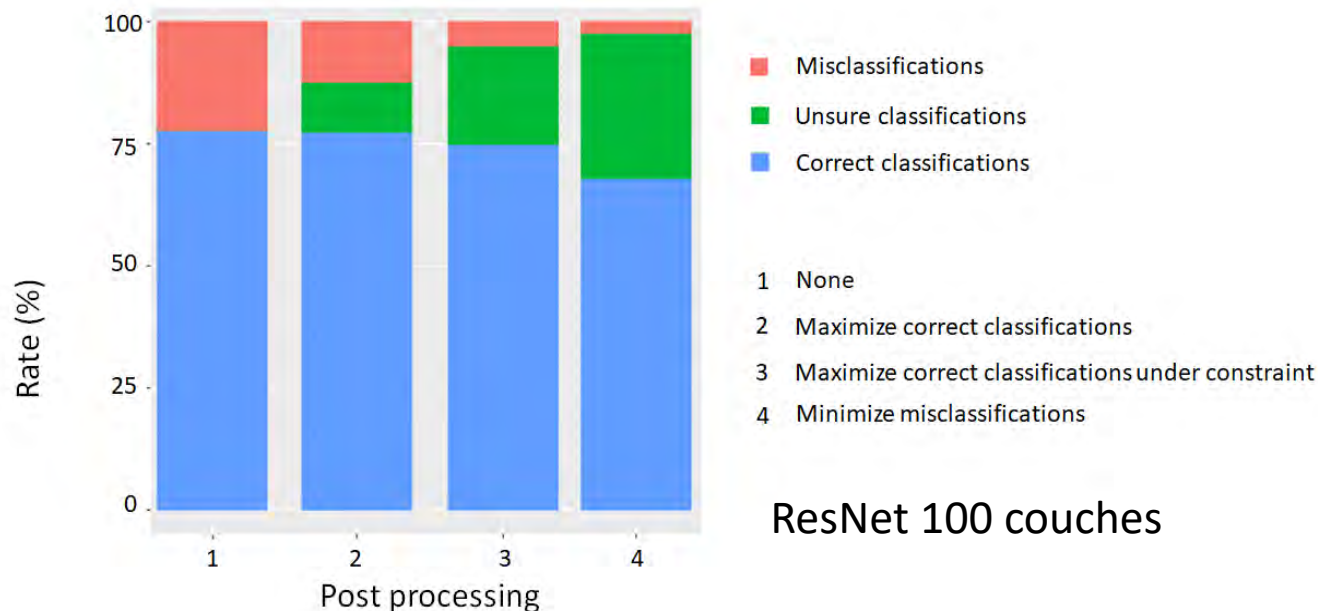
$$\max \left[\frac{\#\{\text{Vrai Positif}\}}{\text{taux classif. correcte}} \mid \left(\frac{\#\{\text{Misclassif}\}}{\text{taux mauvaise classification}} < 5\% \right) \right]$$

- Seuil 3 : erreur de classif minimum

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

Résultats

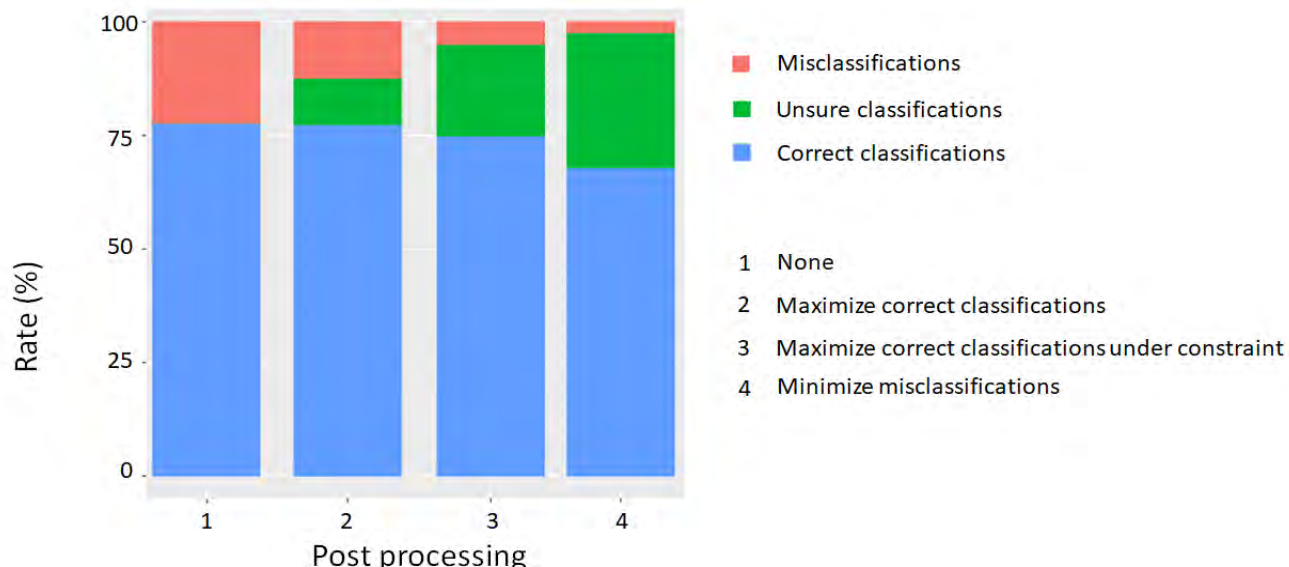
- Data-set (Ile de Mayotte) :
 - D0 : 130 vidéos, 69 169 images, 20 espèces.
Entre 1 134 et 7 345 imageries par espèce.
 - D1 : 20 vidéos, 6 320 images
 - D2 : 25 vidéos, 13 232 images



Les scénarios...

Qui va traiter les « unsure » ?

- CAS1 et CAS2 : Personne
Détection d'événements (peu importe le taux d'erreur)
 - monitoring d'espèces invasives
 - événement rare
- CAS3: Des humains ; les “unsures” sont vraiment “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 :

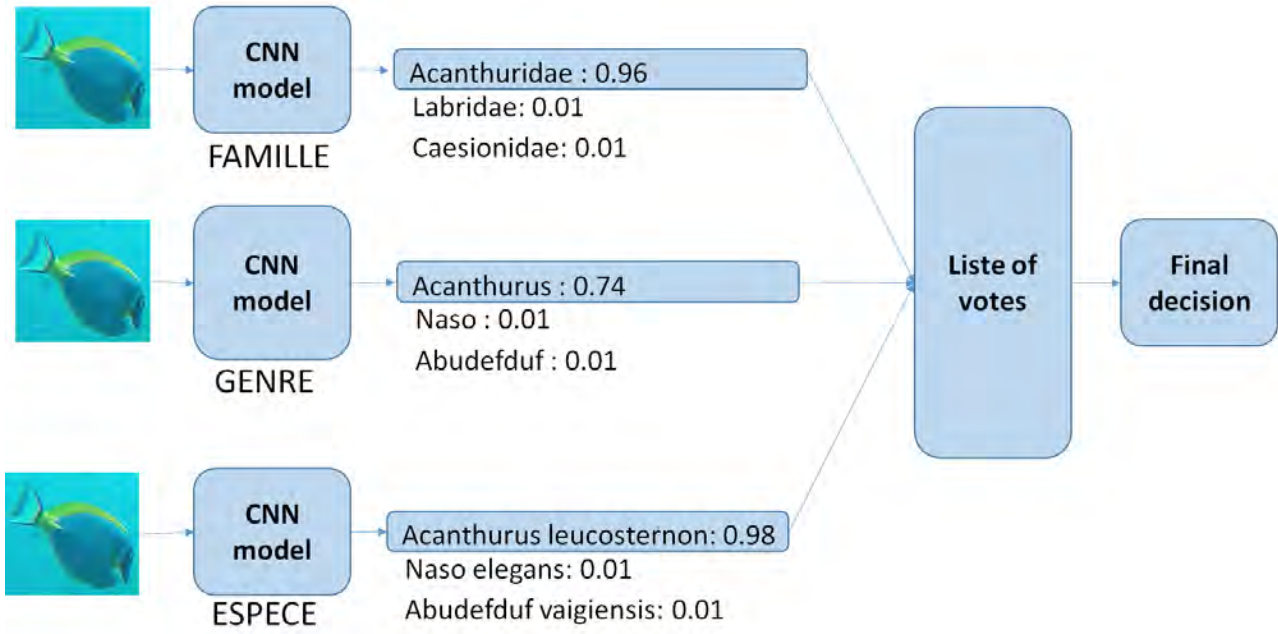
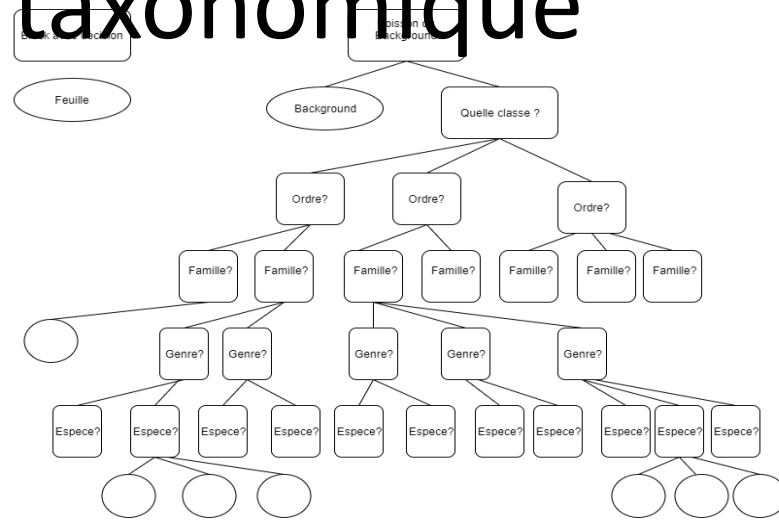
CAS1: Les votes (confiants) sont cohérents

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 !

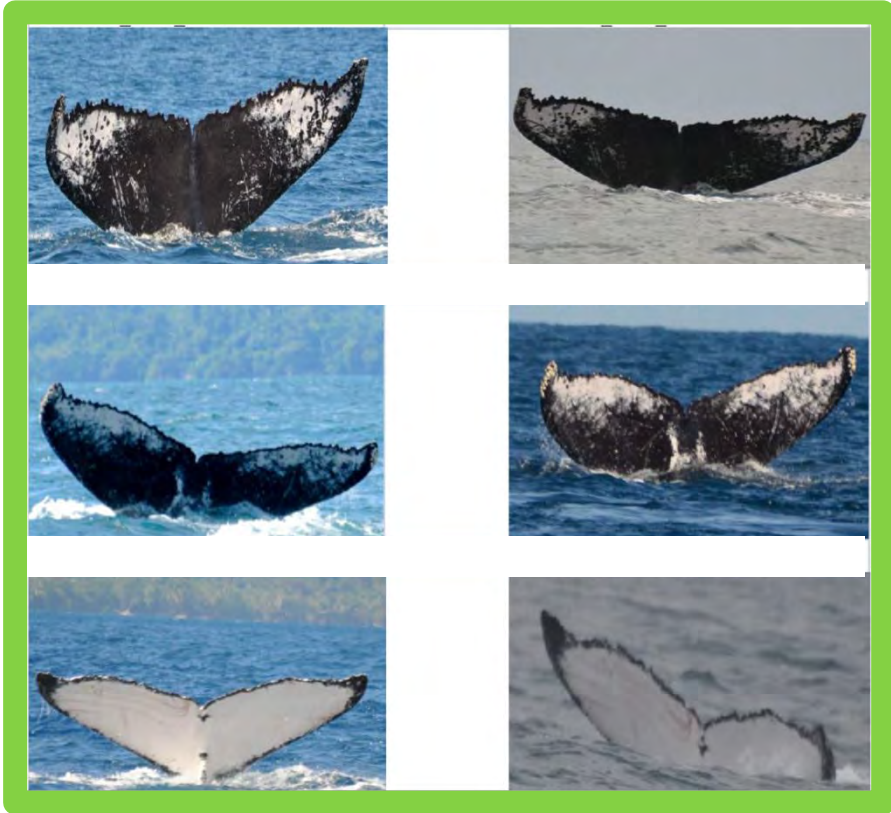
Unsupervised identification

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 spatiale (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 »

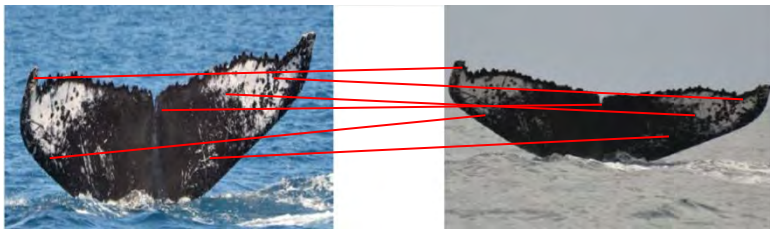
Estimer une transformation géométrique entre l'image à appareiller et l'image cible pour affiner l'apparition

- 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



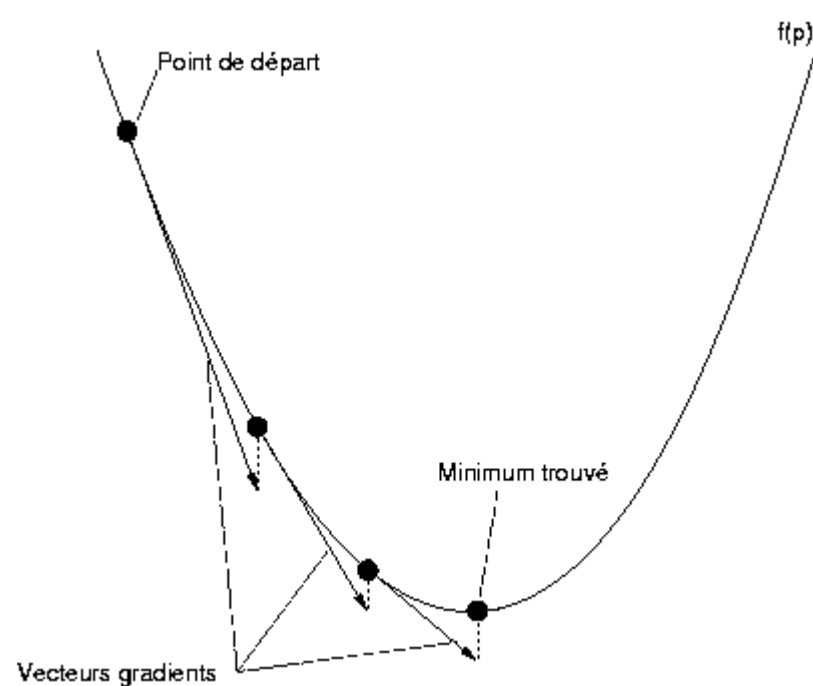
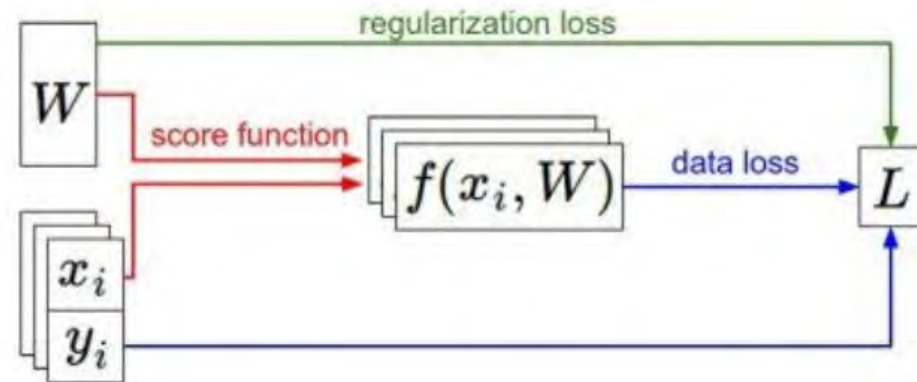
En 1D la direction de la pente est égale au négatif de la dérivé ☺

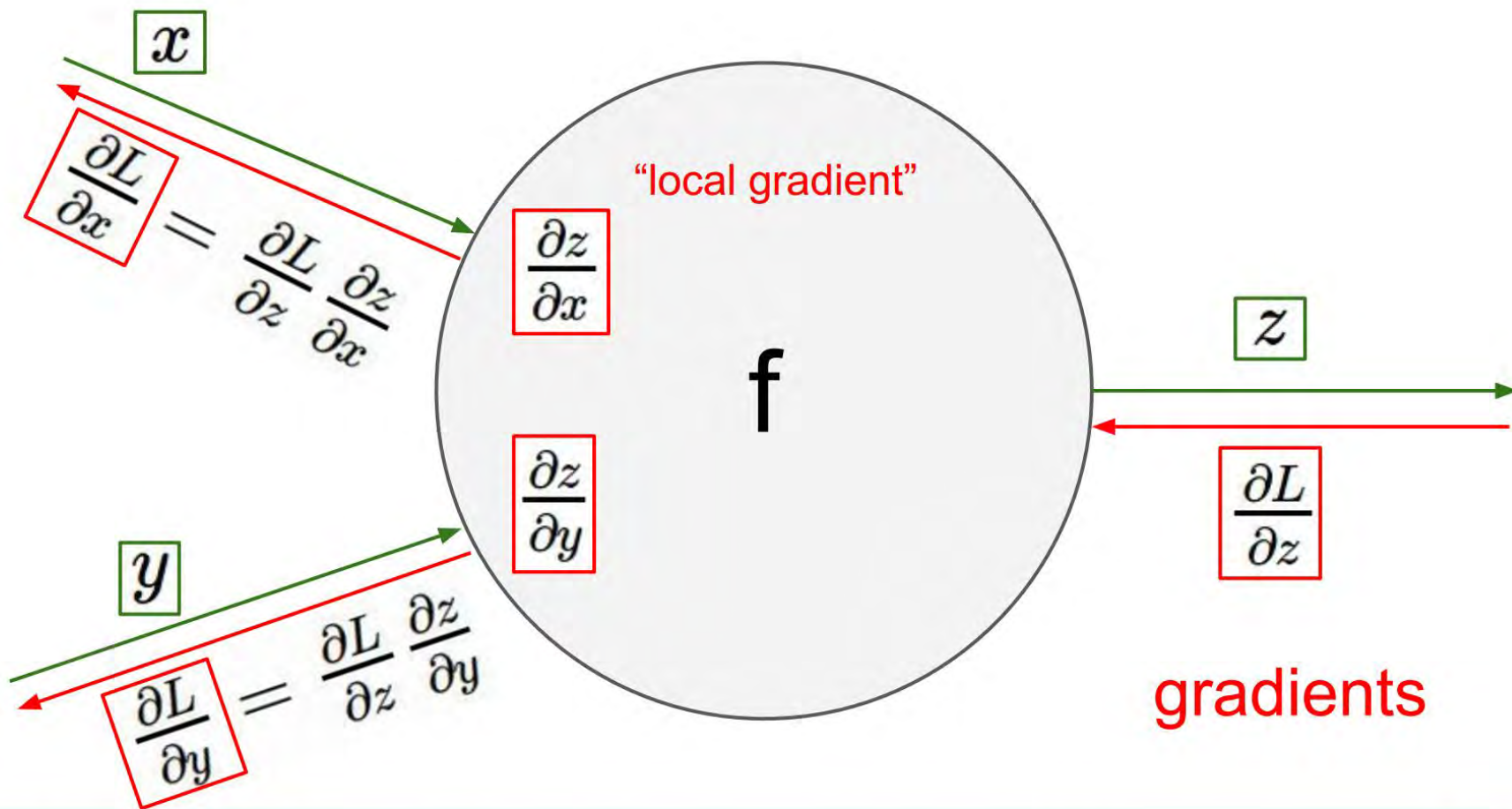
$$\frac{df(x)}{dx} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

$$L_i = -\log\left(\frac{e^{s y_i}}{\sum_j e^{s_j}}\right) \quad \text{Softmax}$$

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

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





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

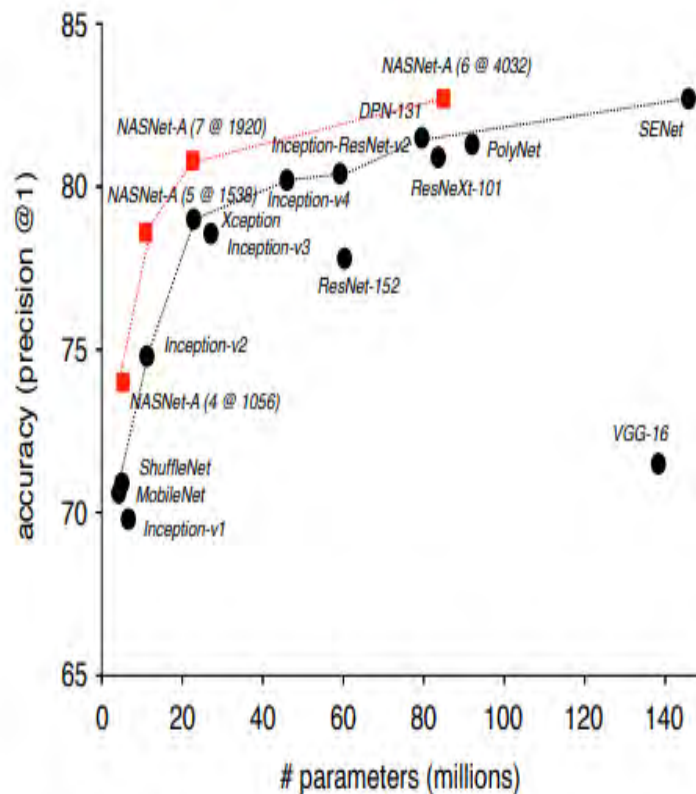
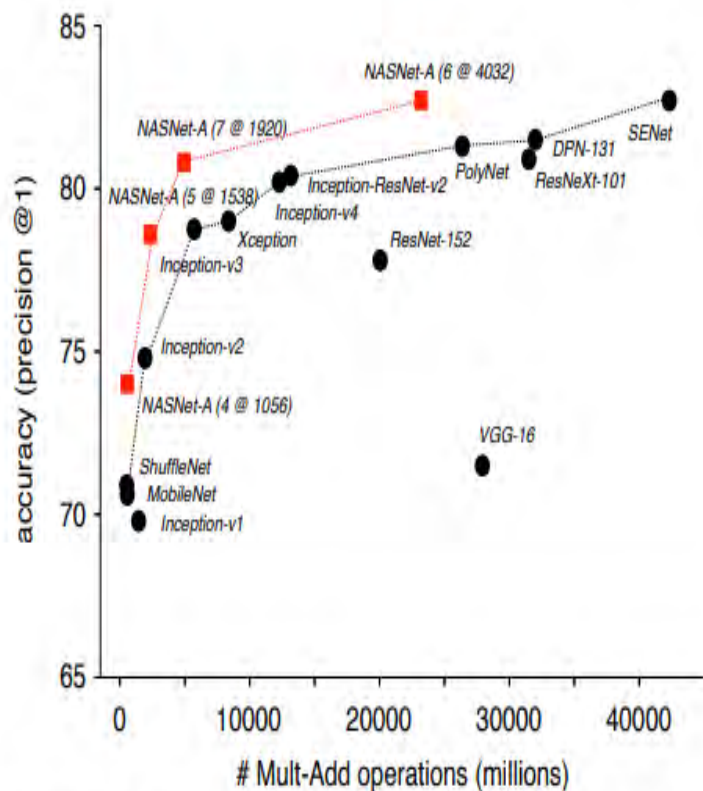


Figure 5. Accuracy versus computational demand (left) and number of parameters (right) across top performing published CNN architectures on ImageNet 2012 ILSVRC challenge prediction task. Computational demand is measured in the number of floating-point multiply-add operations to process a single image.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