Underwater robotics and AI for 3D observation of marine biodiversity

Cyril Barrelet[1], Marc Chaumont[1 3], Gérard Subsol[1], Thomas Claverie[4 5], Sébastien Villeger[4], Houssam Larbi[1], Julian Lilkendey[6], Karen Godary-Dejean[2]

[1] Research-team ICAR, LIRMM, Univ Montpellier, CNRS, Montpellier, France
[2] Research-team EXPLORE, LIRMM, Univ Montpellier, CNRS, Montpellier, France
[3] Univ Nîmes, France
[4] MARBEC, Univ Montpellier, Montpellier, France
[5] CUFR, Centre Universitaire de Formation et de Recherche de Mayotte, France
[6] Data Science Center, University of Bremen

Objectives

1. Automatic fish identification
2. Automatic fish counting

Biodiversity

3. Automatic fish measurement → Biomass

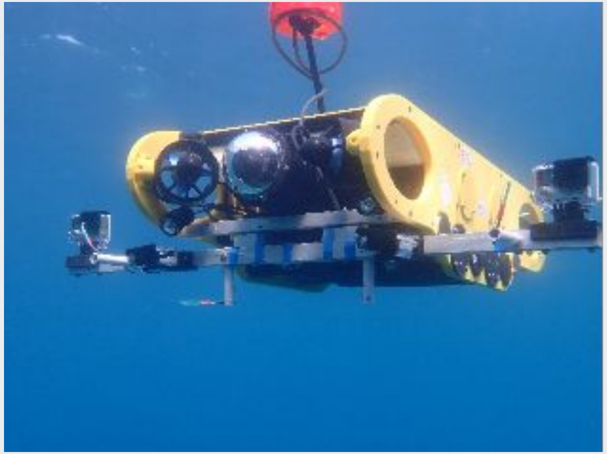
↳ 3D information

4. Automatic assessment of fish behaviour - functional traits

3D data acquisition (1)

Current stereo data acquisition system

Future stereo data acquisition system



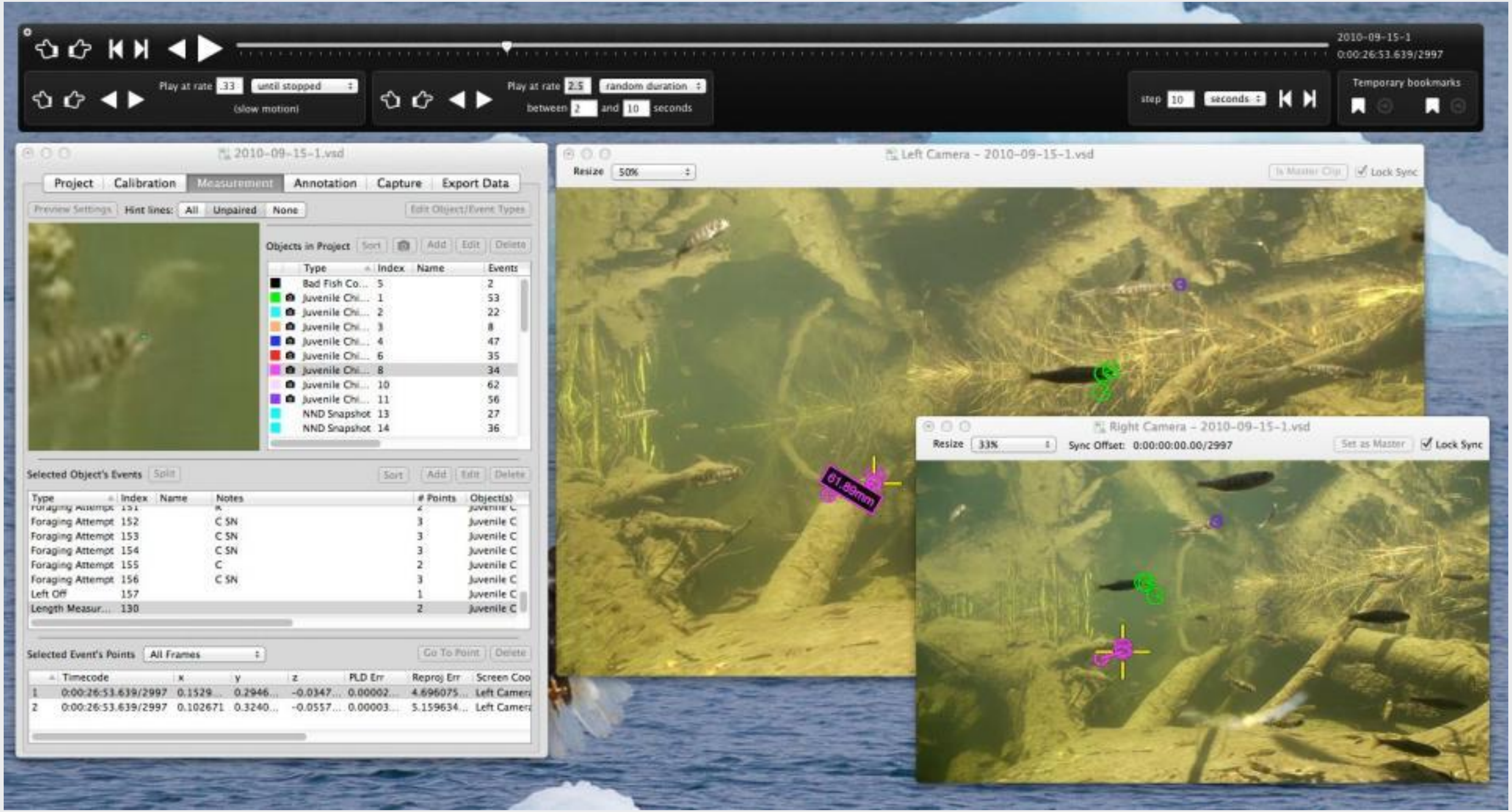
3D data acquisition (2)

- +1,000h of stereo video from MARBEC

Left view :

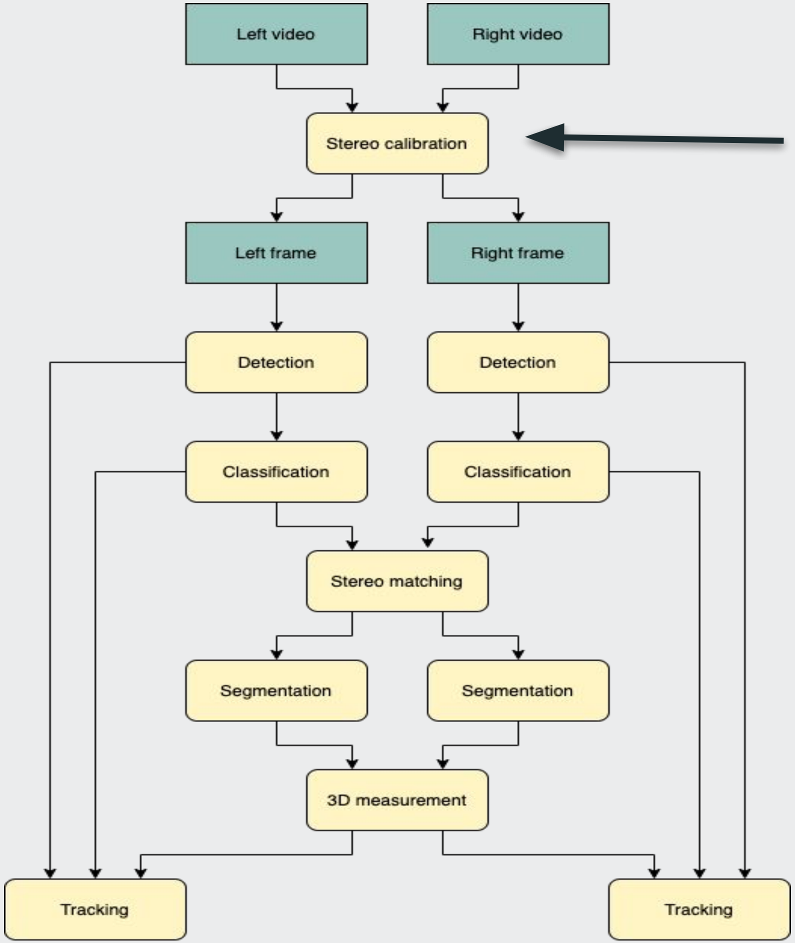
Right view :

State of the art: VidSync[1]

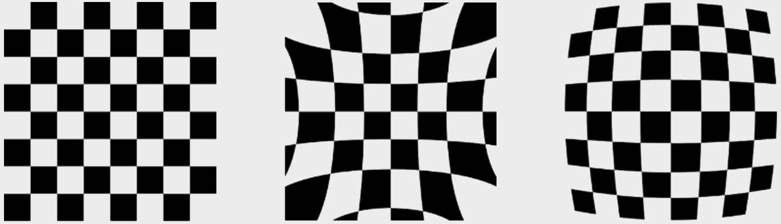


1. http://www.vidsync.org/
“Measuring fish and their physical habitats: versatile 2D and 3D video techniques with user-friendly software”, J R, Neuswanger et al (2016)

Processing pipeline

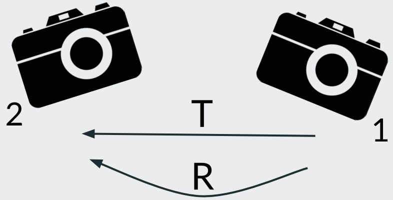


$$
s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\overbrace{\begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}}^{\text{Intrinsic parameters}} \overbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix}}^{\text{Extrinsic parameters}}}_{\text{Projection matrix}} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}
$$



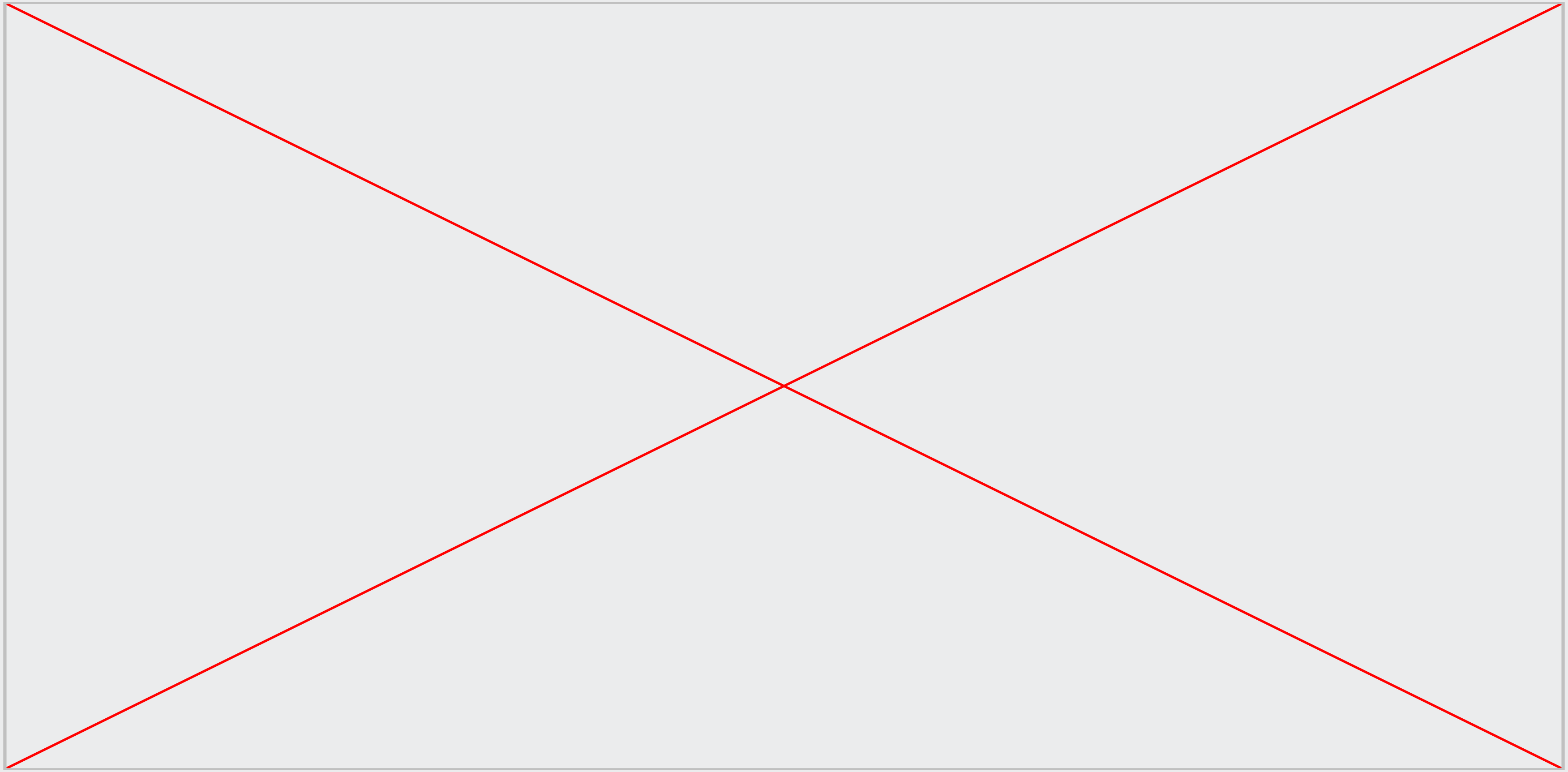
No distortion | **Negative radial distortion** | **Positive radial distortion**

$$
\begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ 1 \end{bmatrix}
$$



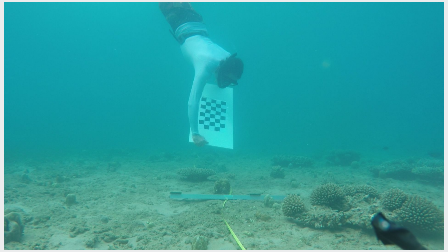
Camera calibration (1)

Left view **Right view**

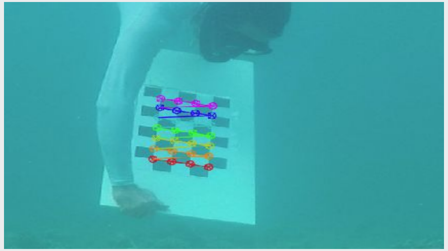


Camera calibration (2)

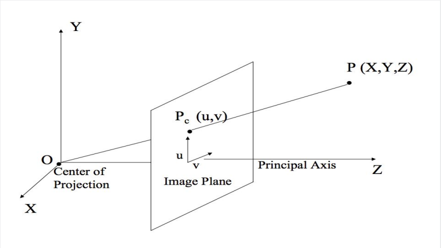
Get intrinsic and extrinsic parameters, and radial distortion coefficients



- Take several pictures of the chessboard



- Detect chessboard corner with Harris corners detector[1]
- Apply Tsai's camera calibration algorithm[2]



1. "A Combined Corner and Edge Detector", C. G. Harris et al (1988)
2. "An efficient and accurate camera calibration technique for 3D machine vision", in IEEE Conference on Computer Vision and Pattern Recognition, J. R. Tsai (1986)

Stereo system calibration (1)

Determine the transformation between the two cameras

- Identify a number of point in the two cameras and compute their relative position and orientation[1]

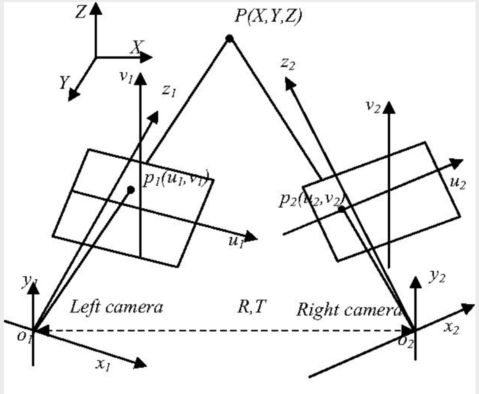
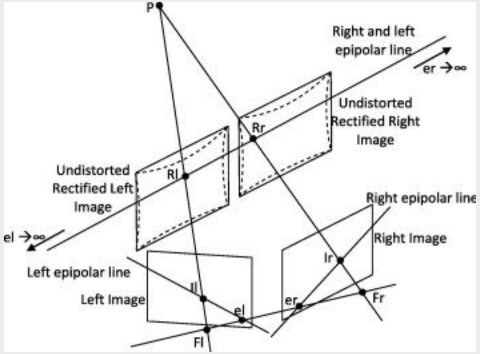


Fig. 1 Stereo Camera Calibration

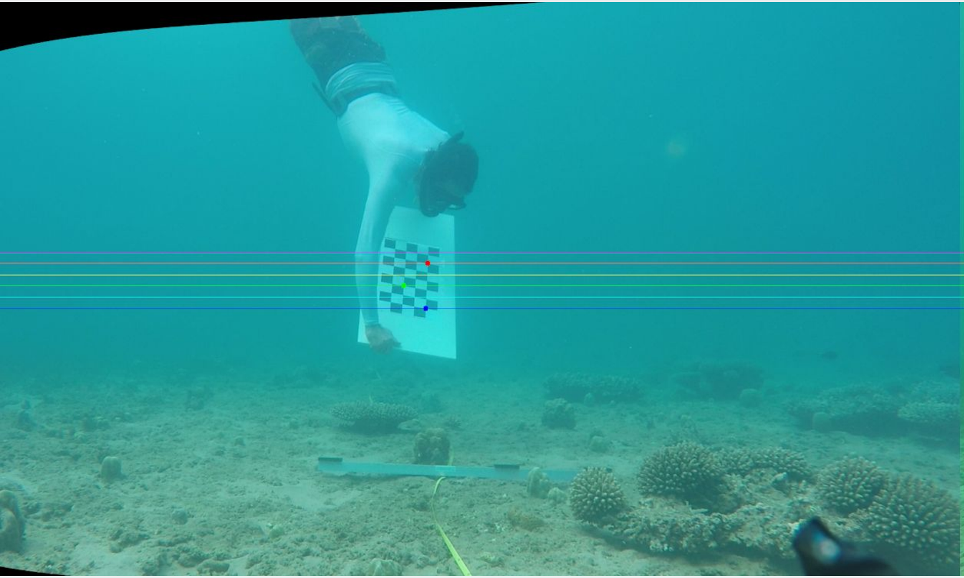
- Rectify the images



1. "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses", J. R. Tsai (1987)

Stereo system calibration (2)

Left view



Right view

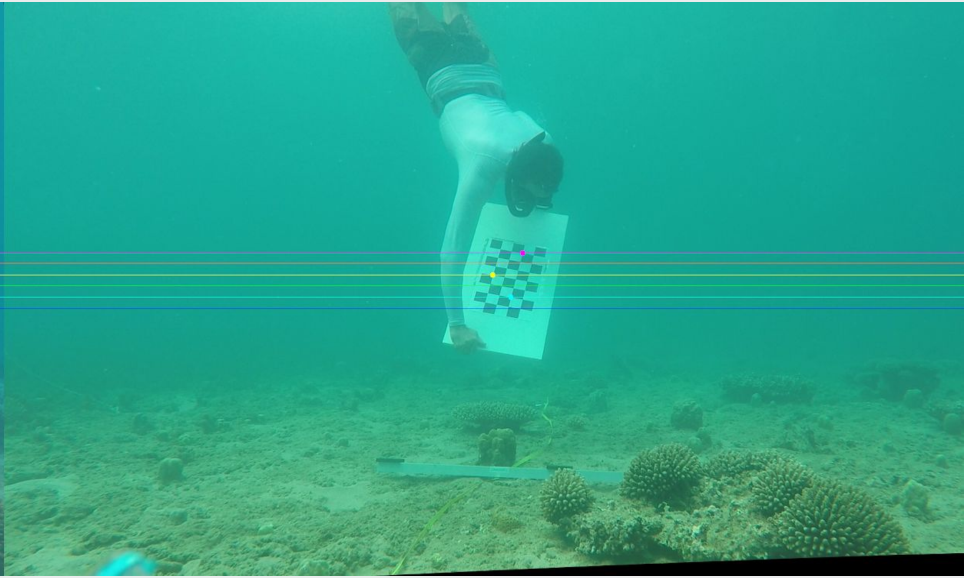
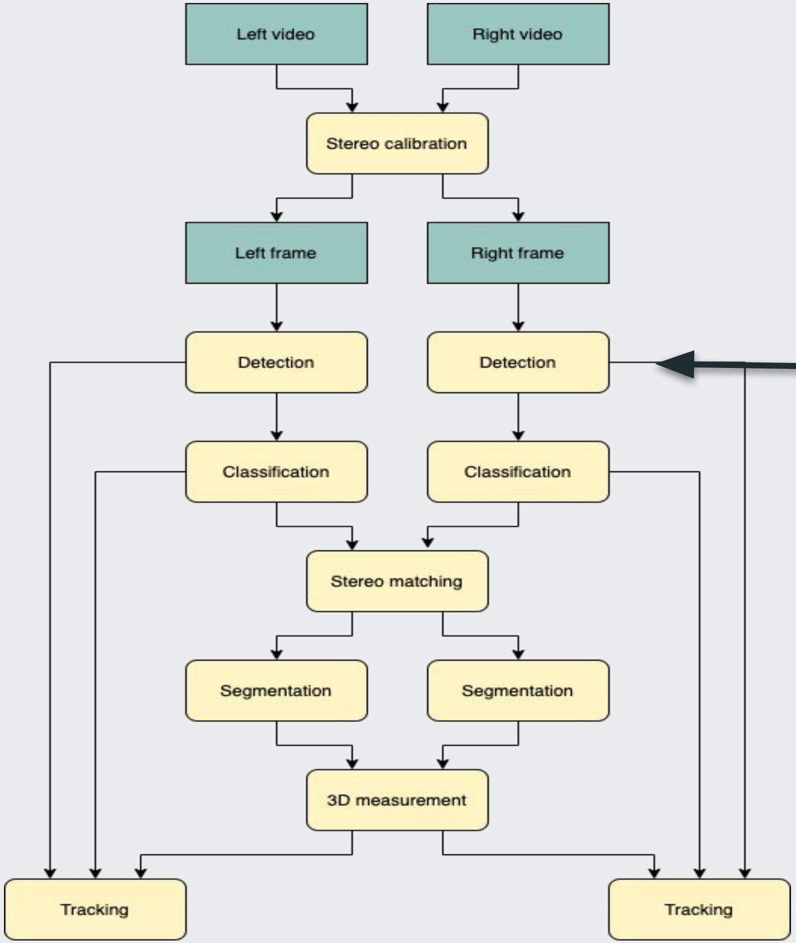
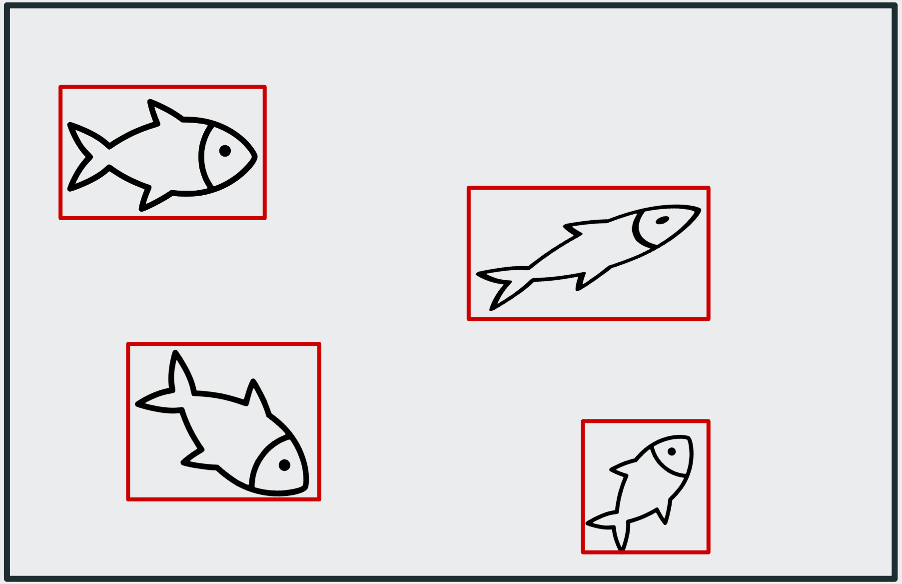


Image rectification

Processing pipeline





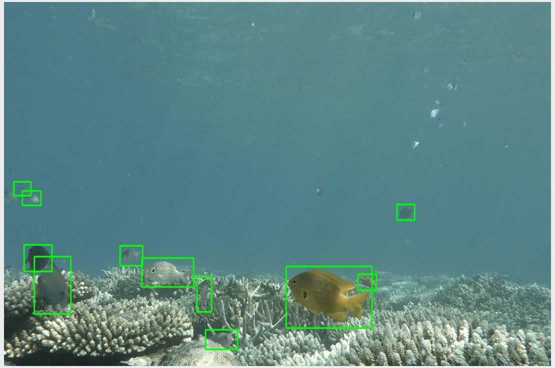
Fish detection (1)

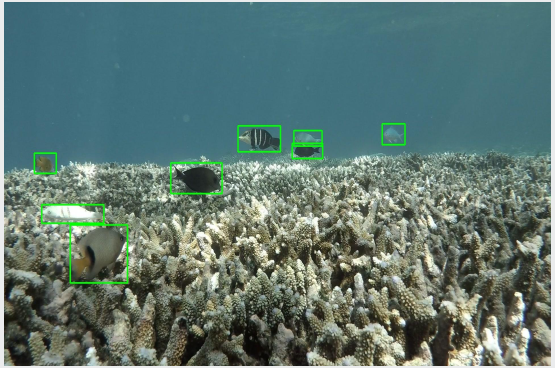
Define a Bounding-Box where a fish is detected

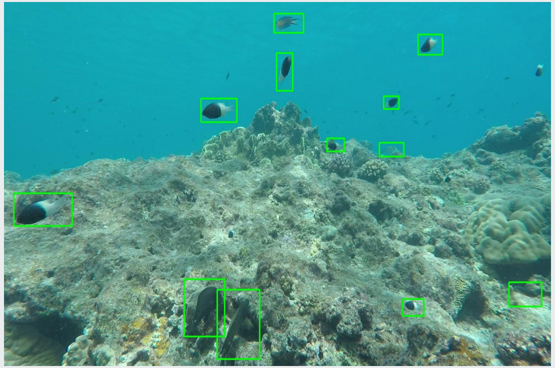
- Create a labelled database by hand

Fish detection database (2)

- 3,362 labelled frames from MARBEC taken in Mayotte, Scattered Islands, Indian Ocean
  - 32,054 bounding boxes (between 0 and 34 fishes per frame)
- 20 classes (species)



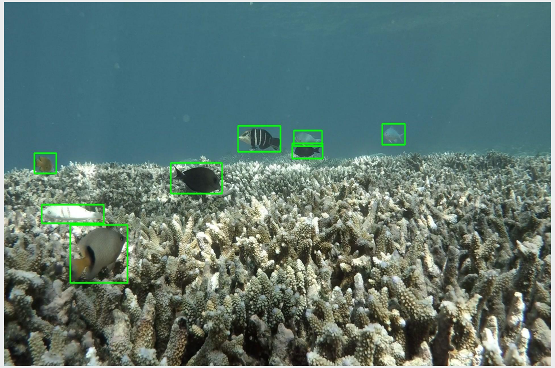




Fish detection (1)

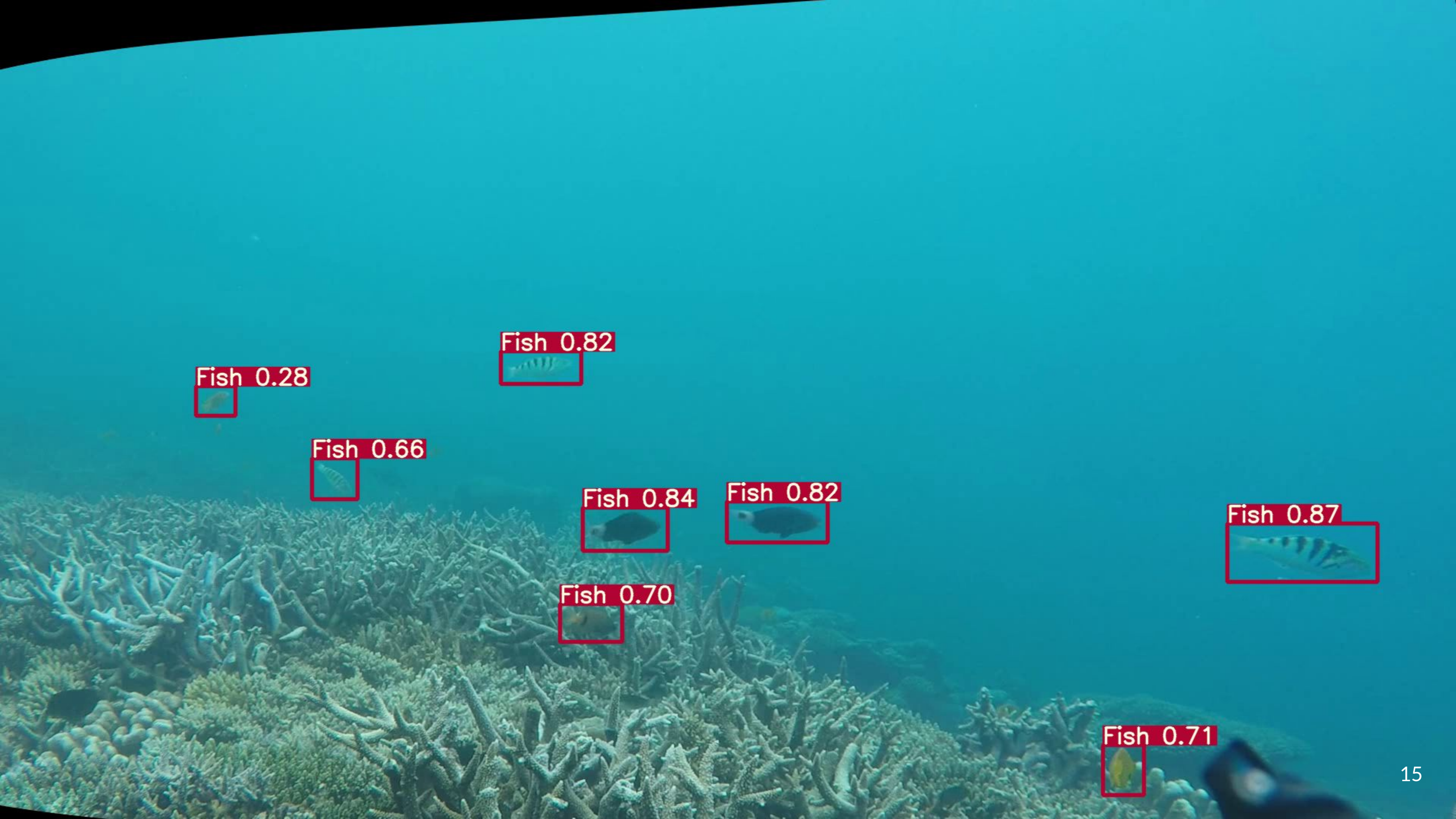
Define a Bounding-Box where a fish is detected

- Create a labelled database by hand
- Augment the database:
  - Random contrast and horizontal flip
- Train a Deep Learning object detection network: YOLOv5[1]

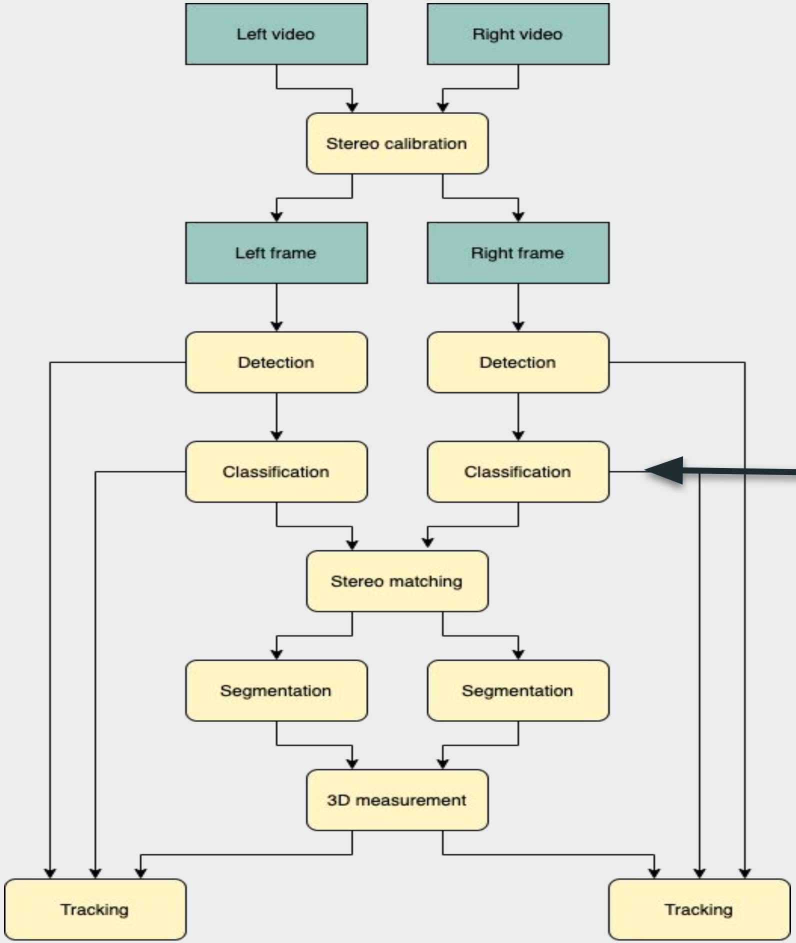


Model	Resolution	Batch size	Precision	Recall	mAP@.5
Yolov5x	640	28	**84.6%**	80.3%	84.8%
YoloV5x6	1280	4	82.9%	**84.7%**	**87.7%**

1. YOLOv5 : https://github.com/ultralytics/yolov5 (2020)


Fish 0.82
Fish 0.28
Fish 0.66
Fish 0.84
Fish 0.82
Fish 0.87
Fish 0.70
Fish 0.71

Processing pipeline



Pomacanthus imperator

Fish Classification (1)

Attribute a species from a species catalog to a Bounding-Box

- Create a labelled database by hand

Fish Classification database (2)

- 44,625 labelled thumbnails from MARBEC
- 64 classes with ≃ 700 thumbnails per class
    - 56 classes for fish species
    - 8 classes without fish (corals, algae, water, etc.)

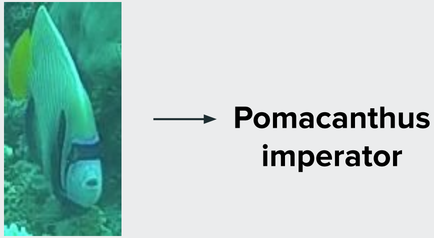
Oxymonacanthus longirostris

Pomacanthus imperator

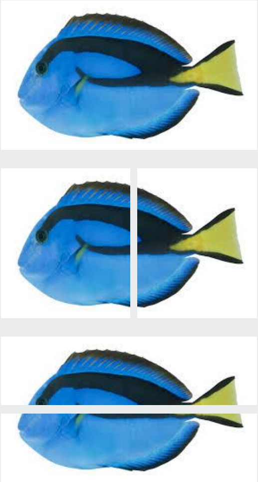
Hard coral

Fish Classification (1)

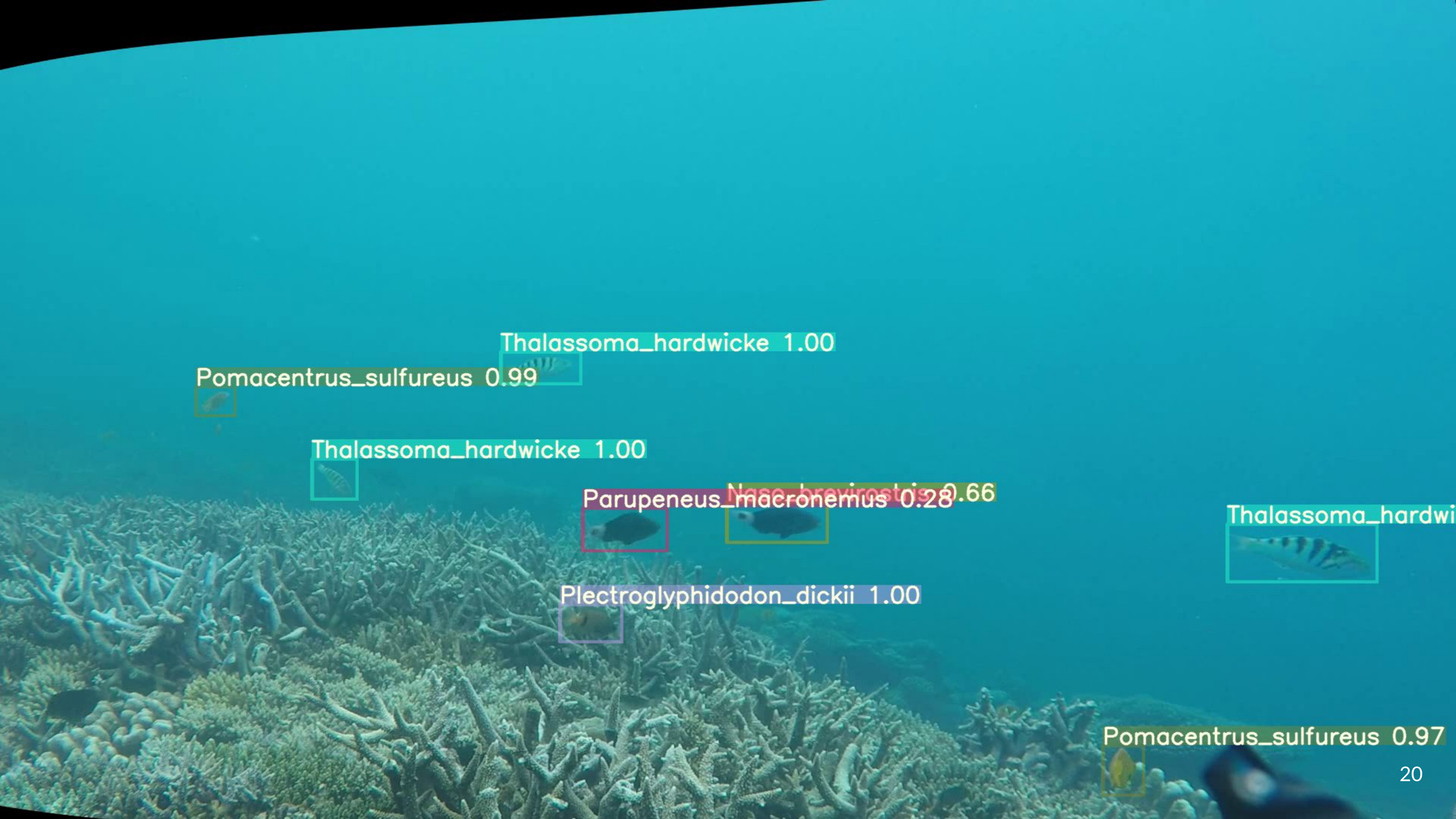
Attribute a species from a species catalog to a Bounding-Box



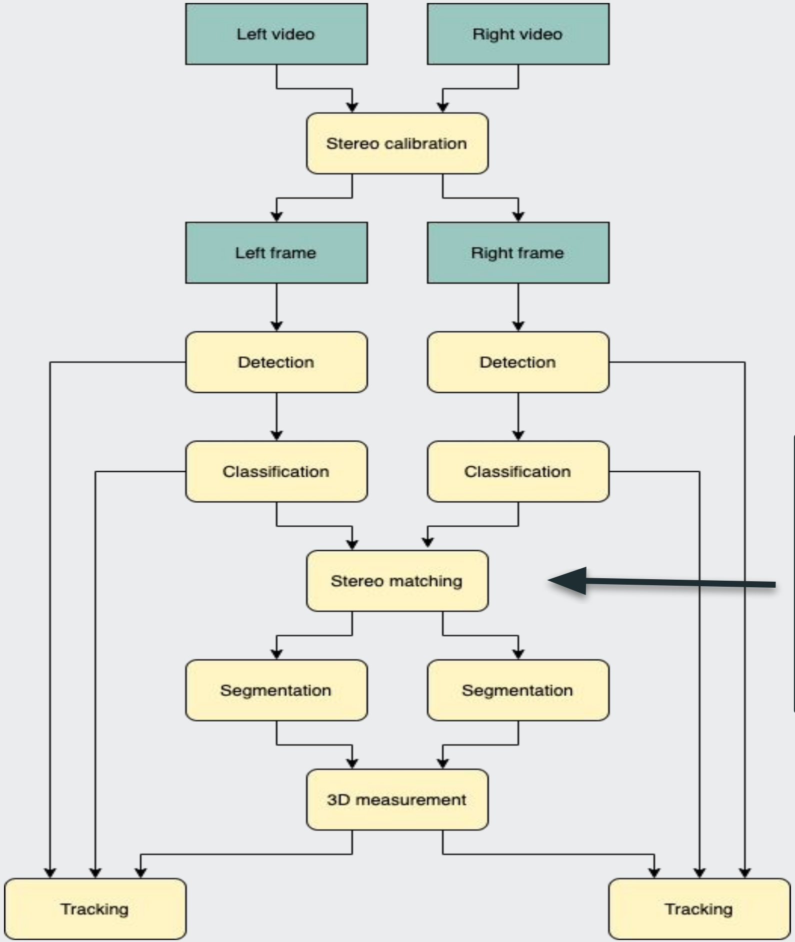
- Create a labelled database by hand
- Augmentation methods:
    - Random contrast and horizontal flip
    - Fish parts cutting[2] (to detect occluded parts)
- 120 classes : 56 species + 56 fishes parts + 8 environments
- Train a deep learning classification network: EfficientNet[1]

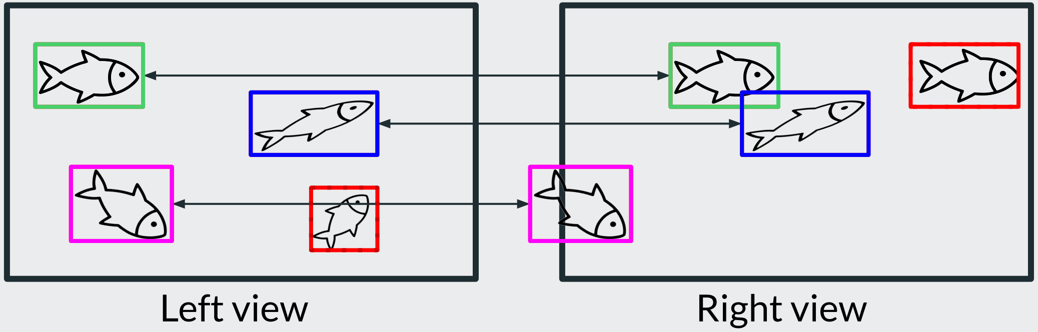


1. "A Deep learning method for accurate and fast identification of coral reef fishes in underwater images", S. Villon, D. Mouillot, M. Chaumont, G. Subsol, T. Claverie, S. Villéger
2. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", Mingxing Tan, Quoc V. Le


Thalassoma_hardwicke 1.00
Pomacentrus_sulfureus 0.99
Thalassoma_hardwicke 1.00
Parupeneus_macronemus 0.28
Thalassoma_hardwi
Plectroglyphidodon_dickii 1.00
Pomacentrus_sulfureus 0.97

Processing pipeline

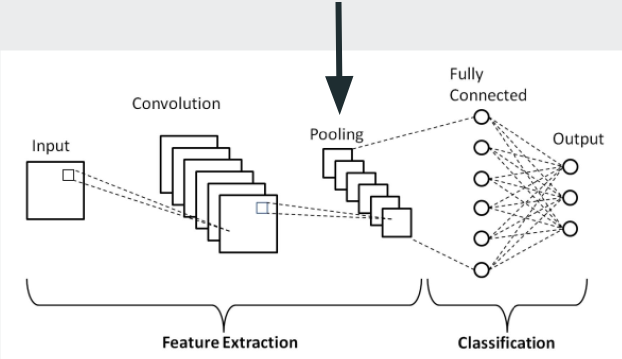




Stereo matching (1)

Find correspondences between the left and right BB

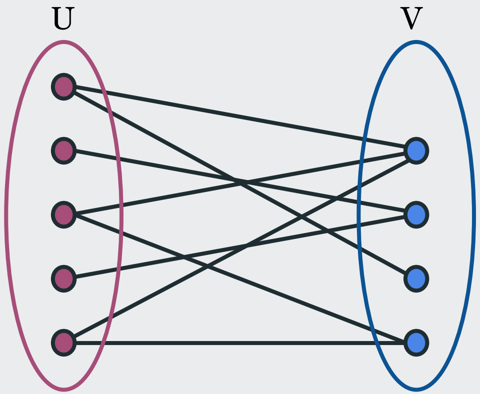
- For each BB on the left and right frame, extract a feature vector from the Global Average Pooling layer of EfficientNet



1

- For each left BBs, look for all the BBs on the right along the same epipolar line, compute distance between their feature vector

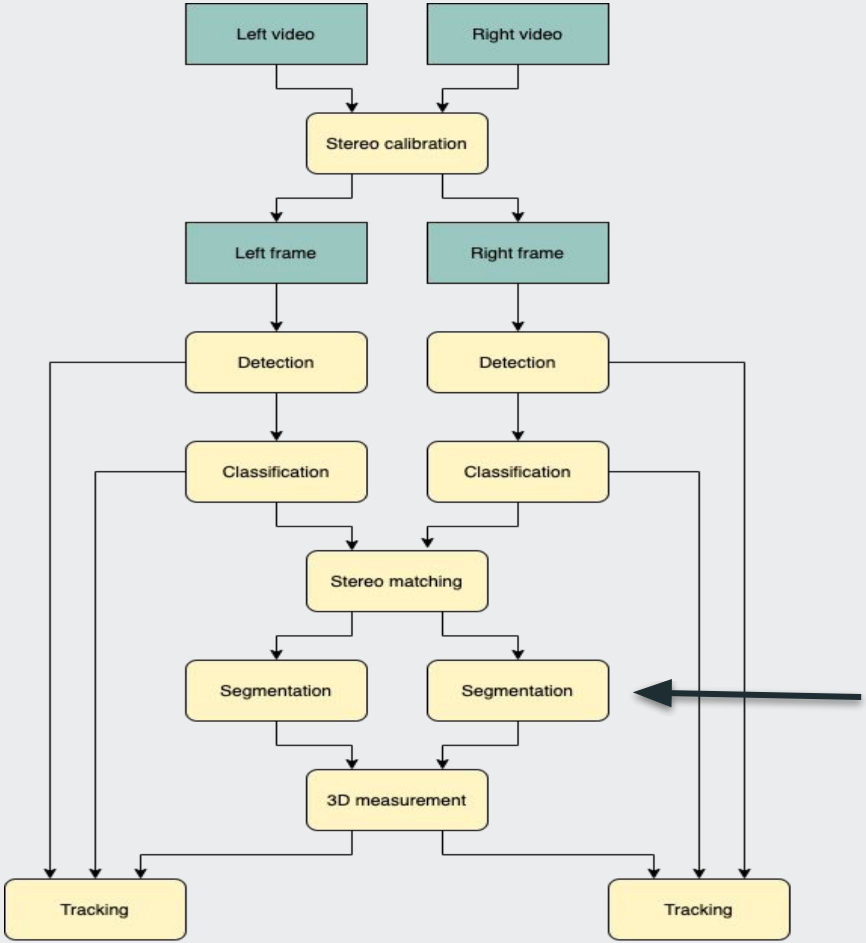
- Solve linear assignment problem using Hungarian algorithm to obtain a one-to-one correspondence



1. “A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets”, V H Phung et al (2019)

Stereo matching (2)

Processing pipeline



Segmentation (1)

Create an accurate mask of the fish in the BB

- Standard image processing algorithms :
  Color thresholding, Canny filter, Snake algorithm, GrabCUT, etc.

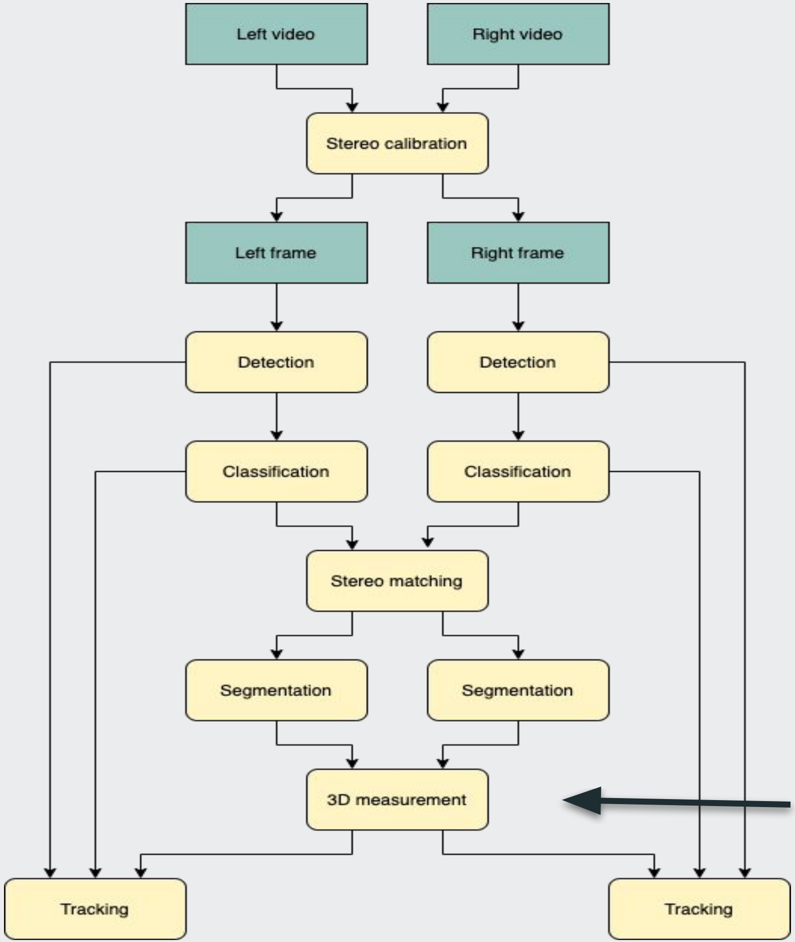


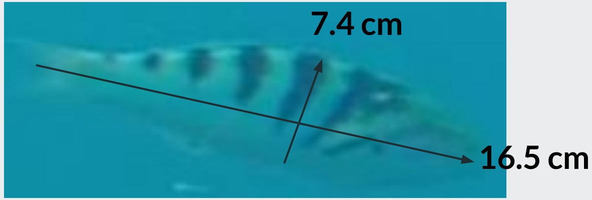
- Create a dataset from the classification database
  - Randomly select 25 thumbnails per species for 1600 thumbnails in total
  - Manually create a mask by forming a polygonal shape
- Train a deep learning segmentation network: DeepLabV3[1]

mIOU
84.4%

1. "Rethinking Atrous Convolution for Semantic Image Segmentation", Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam (2017)

Processing pipeline

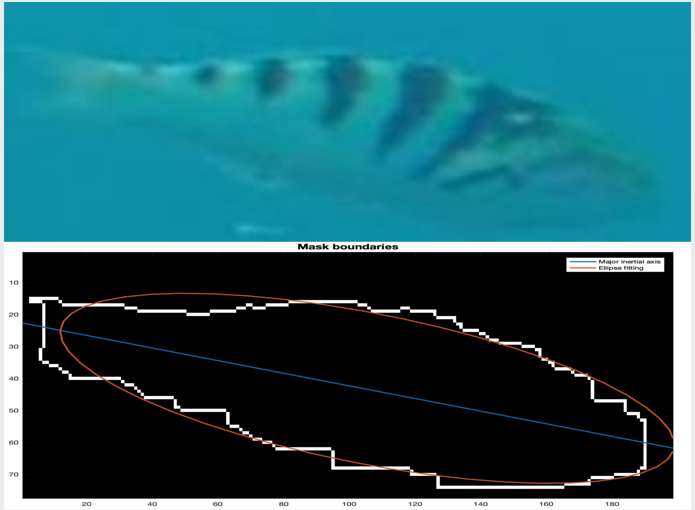




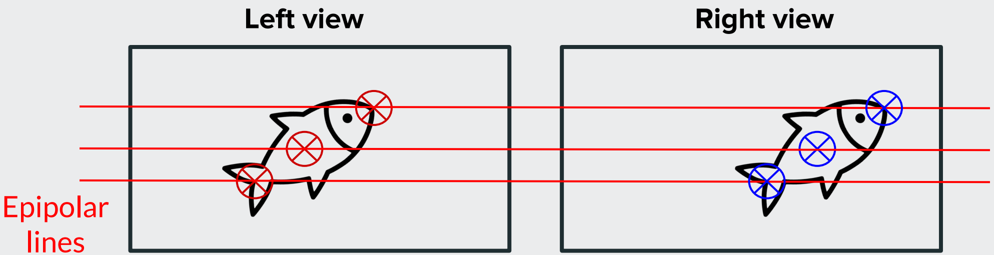
3D measurements (1)

Estimate metric length

- For each BB find extremal and center points by determining the principal axis using PCA



- Find correspondences between these points in left and right BB using epipolar constraint, stereo matching, and image correlation coefficient

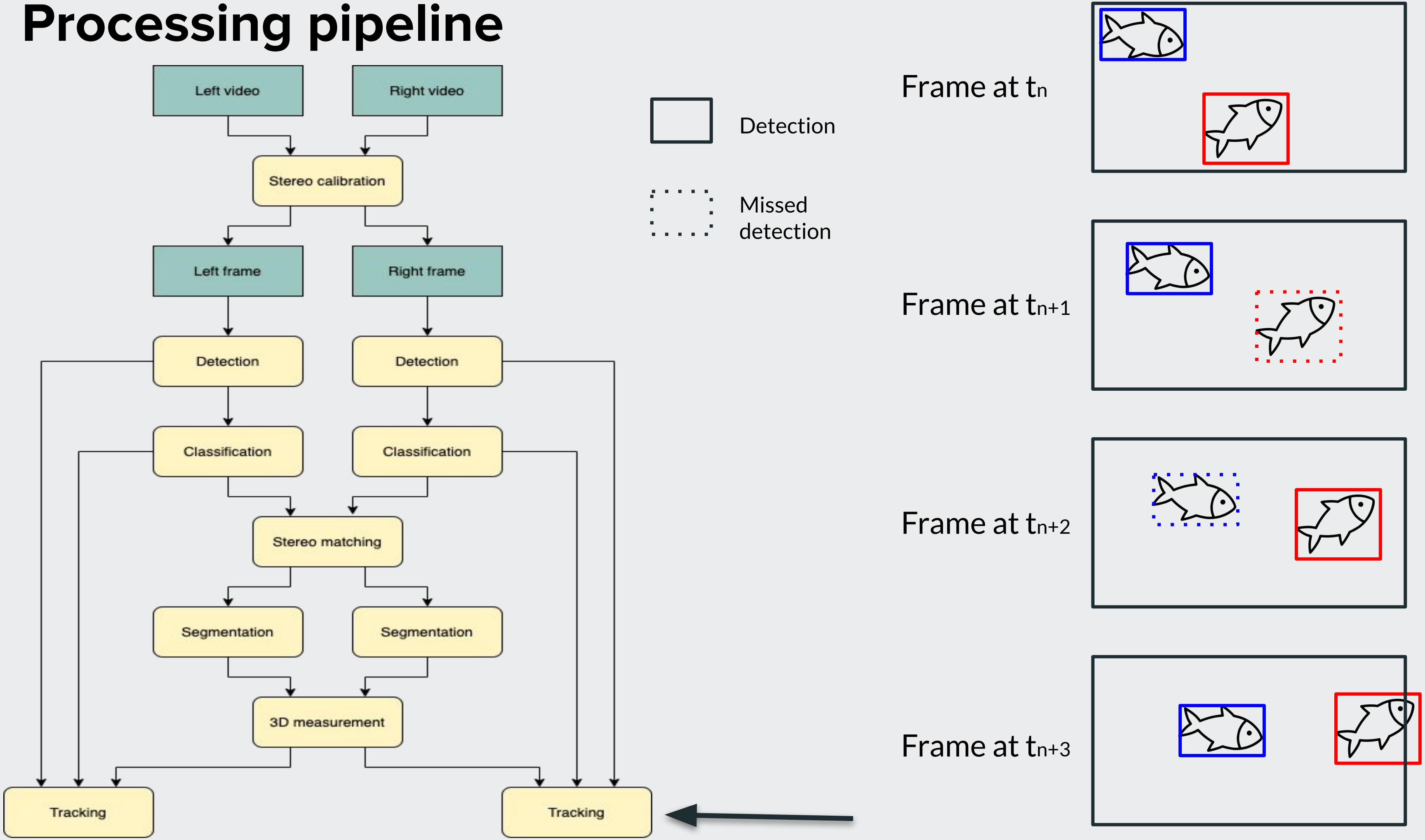


- Compute 3D point coordinates by triangulation
- Estimate the length and the 3D center of the fish

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

3D measurements (2)

Processing pipeline



Tracking (1)

Temporal continuity and consistency

- Tracking algorithm : DeepSORT[1]
  - Motion estimation with a Kalman filter
  - Keep track of the 100 last feature vectors of the same object
  - Associate feature vector similarity and motion estimation
- Allow to “robustify” fish classification and 3D measurements

TOP 5 identification :

- Pomacanthus imperator : 0.68
- Pomacentrus sulfureus : 0.13
- Thalassoma hardwick : 0.04
- Plectro dickii : 0.02
- Parupeneus macronemus : 0.01

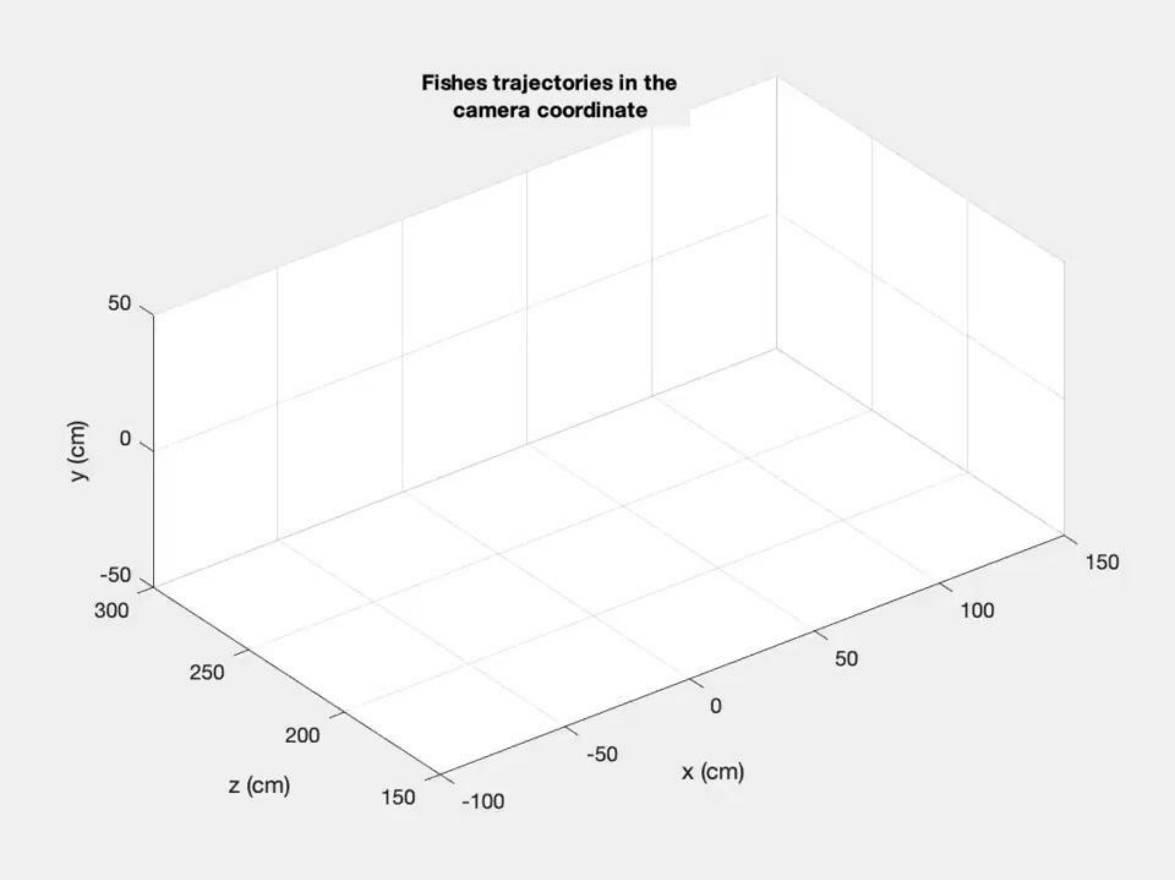
1. “Simple Online and Realtime Tracking with a Deep association metric”, Nicolai Wojke et al (2017)

Tracking (2)


Pomacentrus_sulfureus 0.88
Gomphosus_caeruleus_male 0.37
9.5cm
Pomacentrus_sulfureus 1.00
8.5cm
Plectroglyphidodon_dickii 1.00
5.3cm
Pomacentrus_sulfureus 1.00
7.5cm
Gomphosus_caeruleus
Pom

Towards behaviour analysis

Towards behaviour analysis

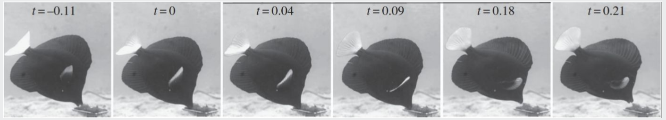


Towards foraging behaviour

Can differences in functional traits (e.g. bites rates and distances between consecutives bites) be explained by the relative energy expenditure for locomotion?

Question: How to know when a fish is foraging?

Hypotheses: A fish does a back-and-forth movement while foraging [1]

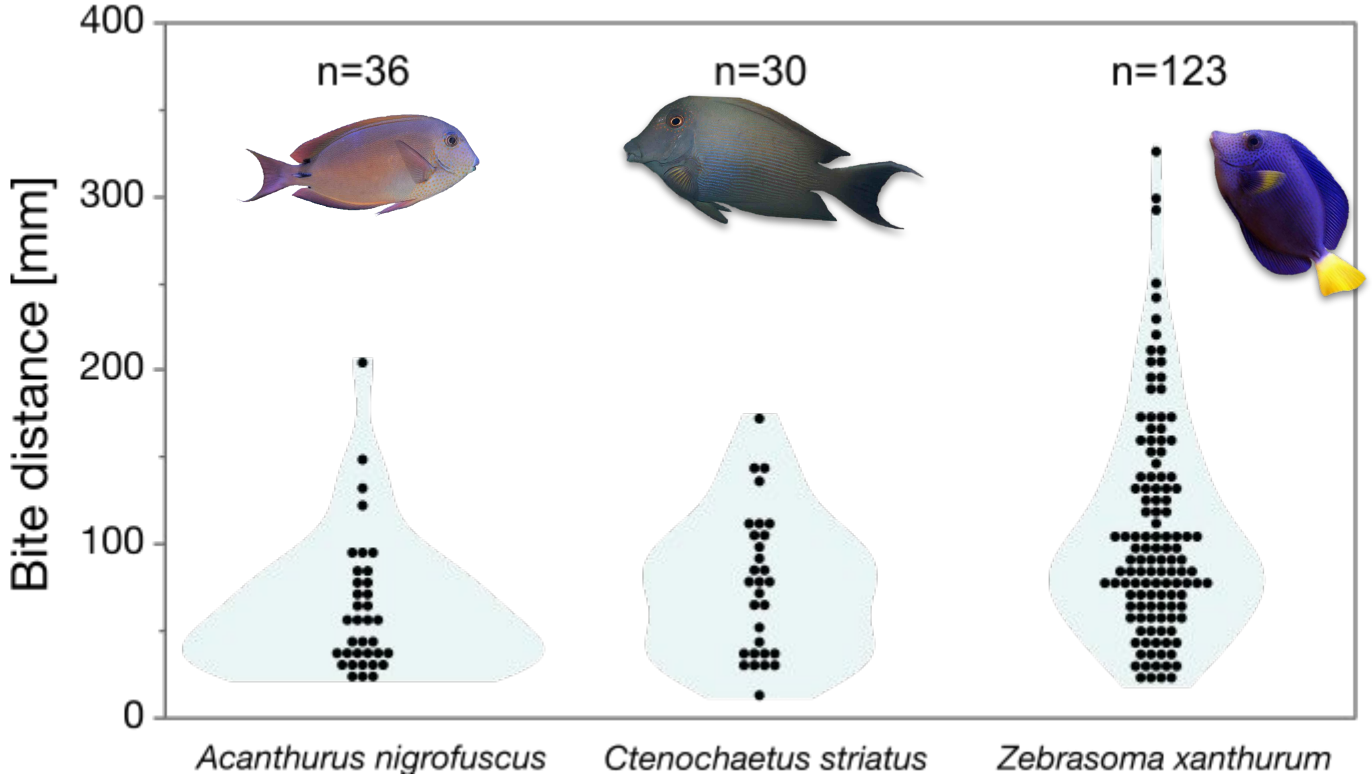


Deep learning issues:

- New ecosystem (Red Sea) → Different domains for detection, classification, and thus, stereo matching
- Train a new classification model on the iNaturalist dataset

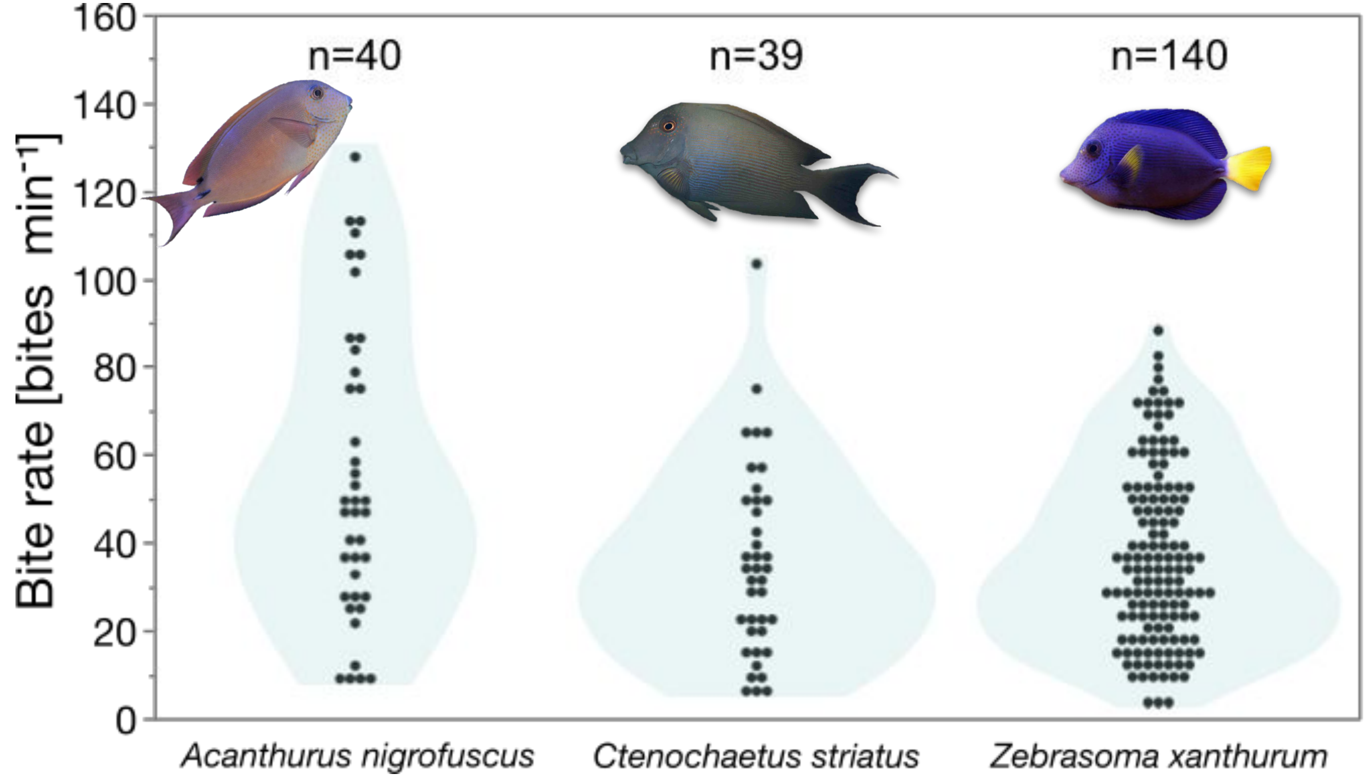
1. “Foraging behaviour in fishes: Perspectives on variance”, B. Marcotte, and H. Browman (1986)
2. https://www.inaturalist.org

Towards foraging behaviour



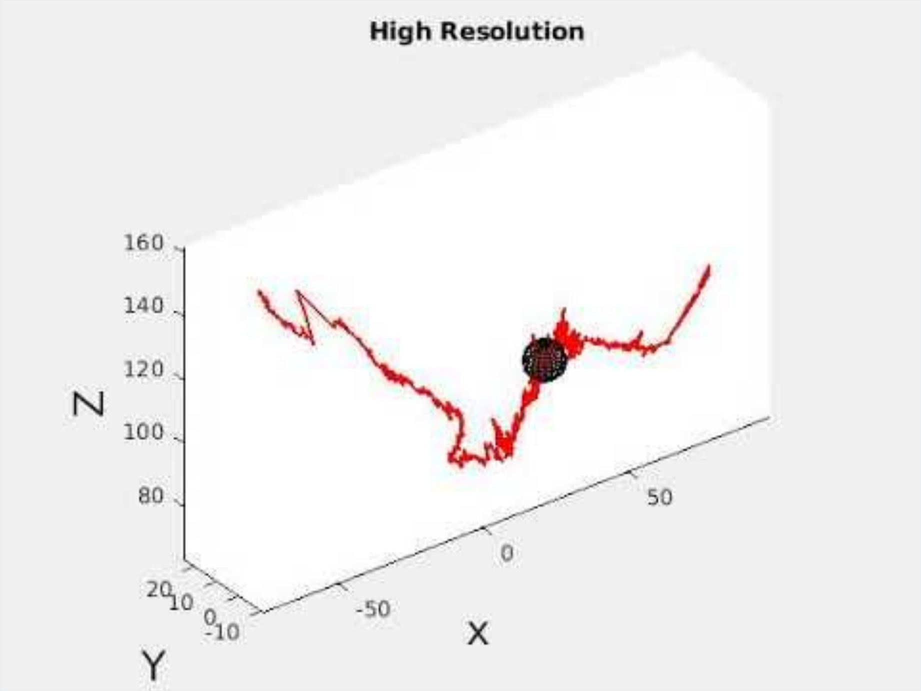
J. Lilkendey

Towards foraging behaviour

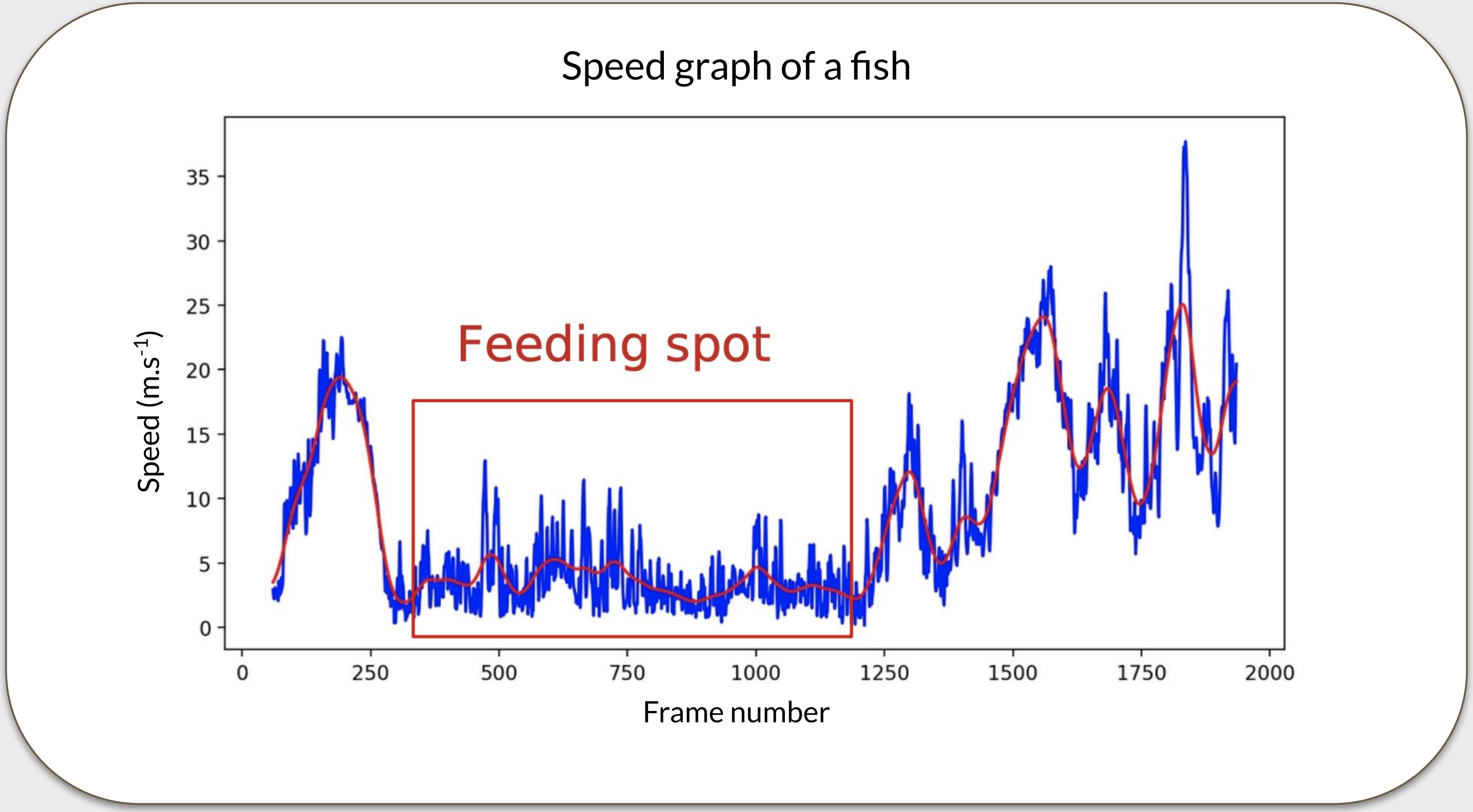


J. Lilkendey

Towards foraging behaviour

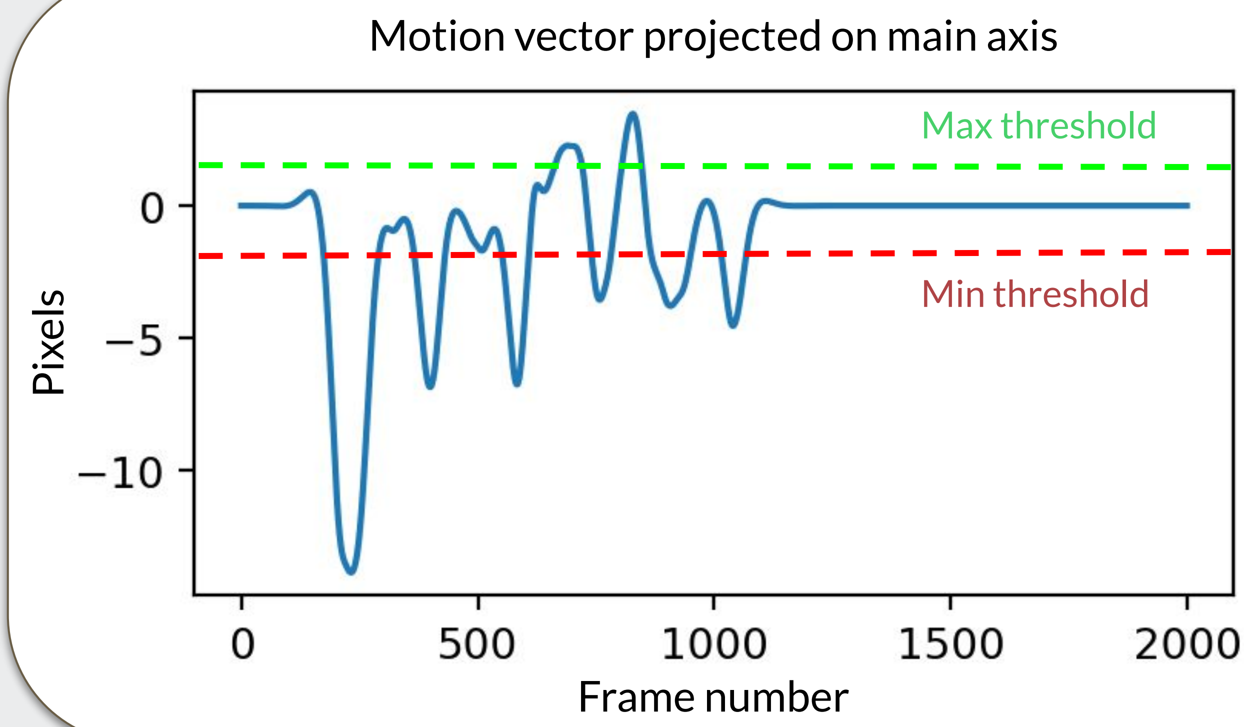


Towards foraging behaviour



Towards foraging behaviour

Towards foraging behaviour



Outlook

- Quick assessment of the biodiversity
  - Automatic fish counting
  - Automatic fish identification
- Quick assessment of the biomass
  - Automatic fish 3D measurement
- Quick assessment of the flow of energy
  - Automatic bites counting
  - Automatic distances measurement between consecutive bites

Underwater robotics and AI for 3D observation of marine biodiversity

Cyril Barrelet[1], Marc Chaumont[1 3], Gérard Subsol[1], Thomas Claverie[4 5], Sébastien Villeger[4], Houssam Larbi[1], Julian Lilkendey[6], Karen Godary-Dejean[2]

[1] Research-team ICAR, LIRMM, Univ Montpellier, CNRS, Montpellier, France
[2] Research-team EXPLORE, LIRMM, Univ Montpellier, CNRS, Montpellier, France
[3] Univ Nîmes, France
[4] MARBEC, Univ Montpellier, Montpellier, France
[5] CUFR, Centre Universitaire de Formation et de Recherche de Mayotte, France
[6] Data Science Center, University of Bremen

An open issue

- An image will always be assigned to class according to the highest score

What should we do when a detected object does not belong to the identification database ?[1]

Classes	δ threshold
0	δ_0
1	δ_1
...	...
n	δ_n

Class	Score
42	0.55

Score > δ_{42} ? → **Correct classification**

Score < δ_{42} ? → **Unsure**

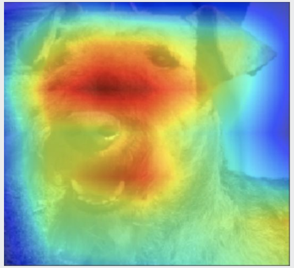
1. “A new method to control error rates in automated species identification with deep learning algorithms”, S. Villon, D. Mouillot, M. Chaumont, G. Subsol, T. Claverie, S. Villéger (2019)

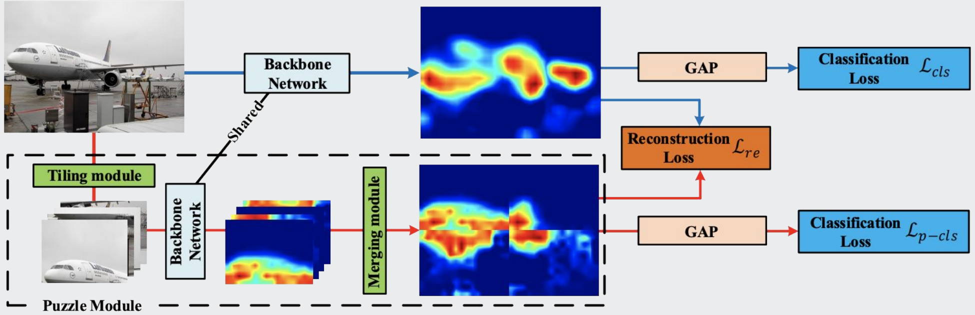
Open issue

How to augment the training database without time-consuming data annotation task? [1]

➜ Weakly supervised segmentation network : Puzzle-CAM[1]

Airedale terrier





1. “Puzzle-CAM: Improved Localization via Matching Partial and Full Features”, Sanghyun Jo, In-Jae Yu (2021)