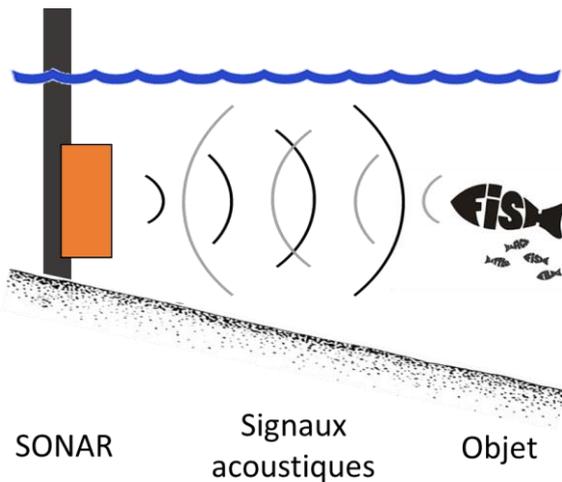


Un réseau de neurones profonds pour détecter et identifier les espèces de poissons sur les données des caméras acoustiques

François MARTIGNAC

Guglielmo FERNANDEZ GARCIA

• L'hydroacoustique pour suivre les populations de poissons



- Non intrusif
- Transposable
- Peu dépendant des conditions environnementales (turbidité...)
- Acquisition haute fréquence de chroniques longues et continues

- MAIS**
- Pas d'accès à l'individu
 - Gestion des données
 - **Identification spécifique**

Martignac et al., 2015

Single-beam Dual-beam Split-beam Multi-beam

1970 1980 1990 2000 2010

Sonars multifaisceaux à très haute fréquences
= **caméras acoustiques**

• L'hydroacoustique pour suivre les populations de poissons



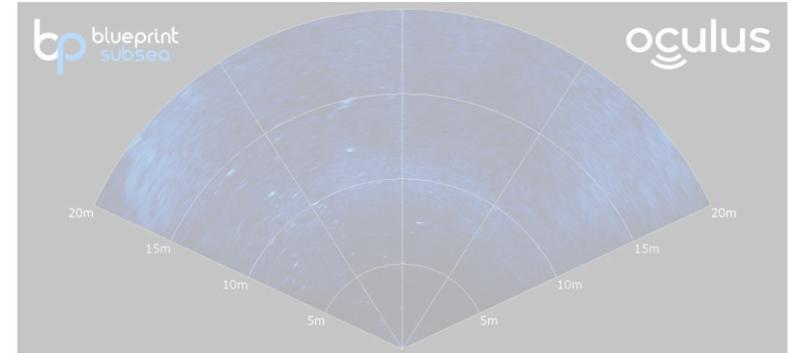
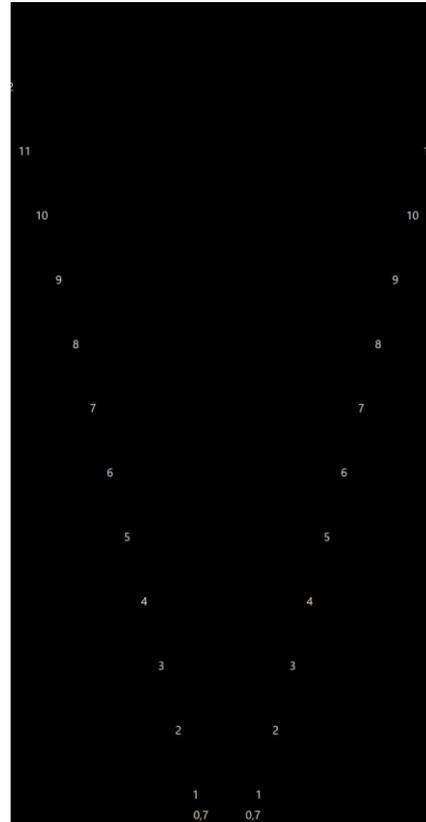
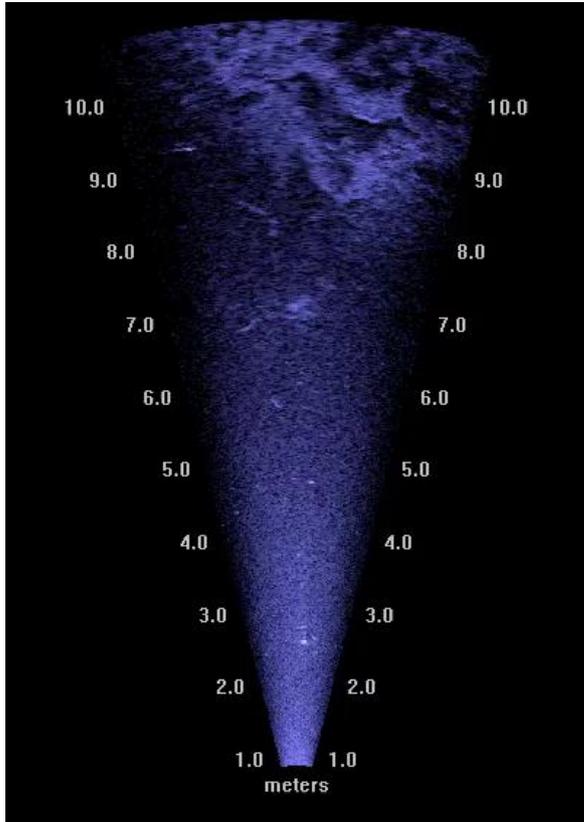
DIDSON (Sound Metrics Corp., US)



ARIS (Sound Metrics Corp., US)



Oculus (Blueprint Ltd., UK)

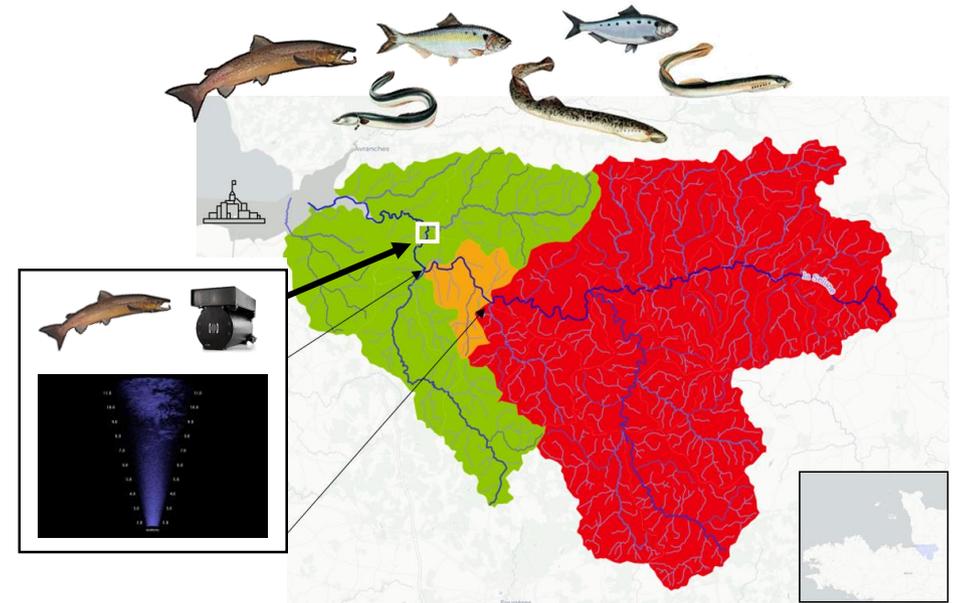


BlueView (Teledyne Inc., US)



• Suivi scientifique de la restauration de la continuité écologique de la Sélune

- Enlèvement de deux barrages (2018 et 2022)
- Suivi des populations de poissons diadromes avant et après les travaux
- Enregistrements continus par hydroacoustique depuis août 2013
 - DIDSON, en HF (1800 kHz), ARIS (ponctuellement)
 - Plus de 100 To de données archivées et partiellement exploitées (saumons majoritairement)
 - Plusieurs milliers de séquences labellisées,
 - Des heures et des heures de vidéos diverses et variées ...





... mais aucune méthode automatique optimale d'analyse des données !



Objectif : automatiser l'analyse des données des caméras acoustiques

- Détecter les passages de poissons
- Identifier les espèces d'intérêt
- Mettre à disposition un outil open-source aux utilisateurs des caméras acoustiques = forte demande internationale*

➤ **Méthodes d'intelligence artificielle : le deep-learning ?**



2019-2021. Post-doc de G. Fernandez Garcia : « Automatisation de l'analyse de données des caméras acoustiques par réseaux de neurones convolutifs (CNN) », stages Master M. Hersant (2021), A. Lebeaud (2022), stage DUT L. Blivet (2021)



* Workshop ANACADIR, mars 2022, Rennes

- Principe du Deep Learning (DL)



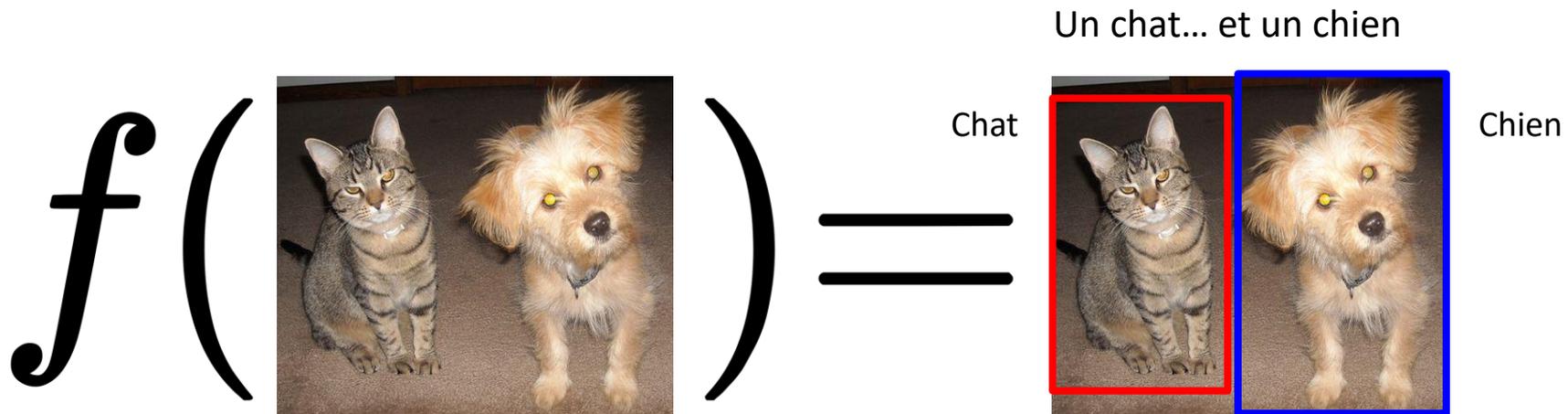
Un chat... et un chien

Chat

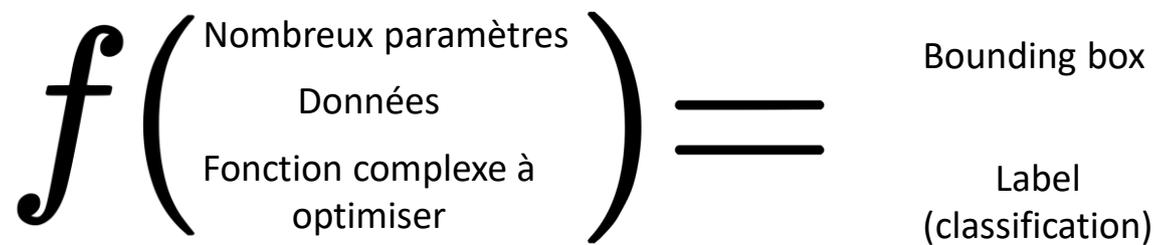


Chien

- Principe du Deep Learning (DL)

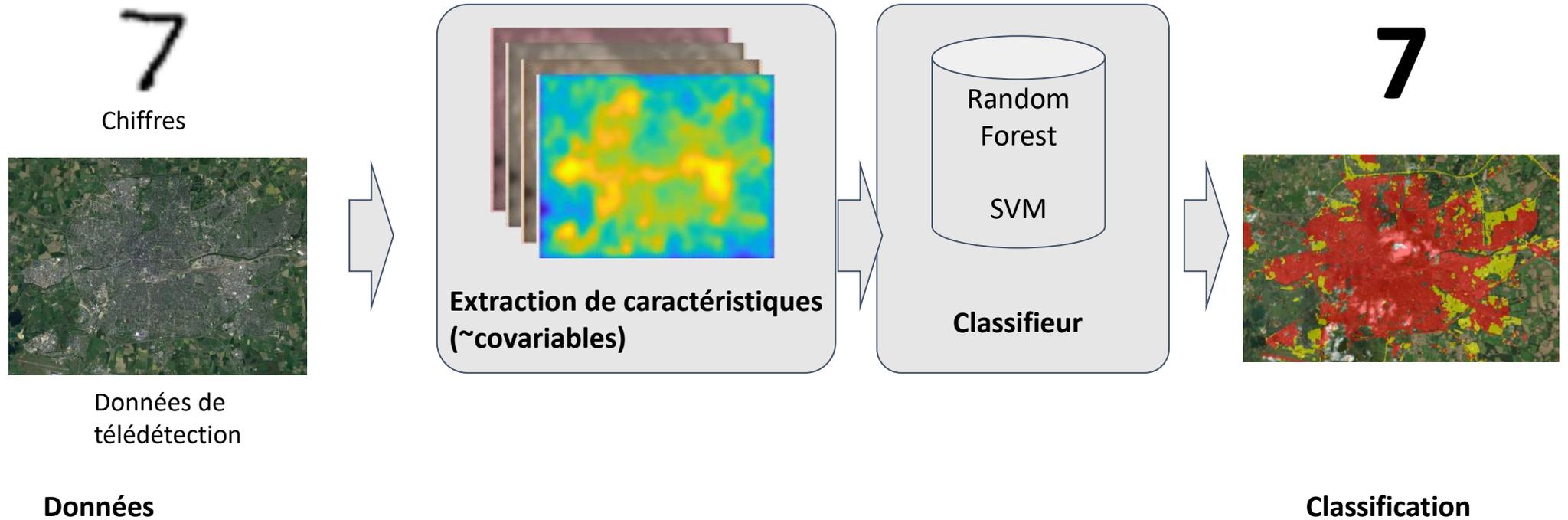


La relation peut être complexe (et non linéaire) ... mais elle existe → l'œil humain peut les différencier !



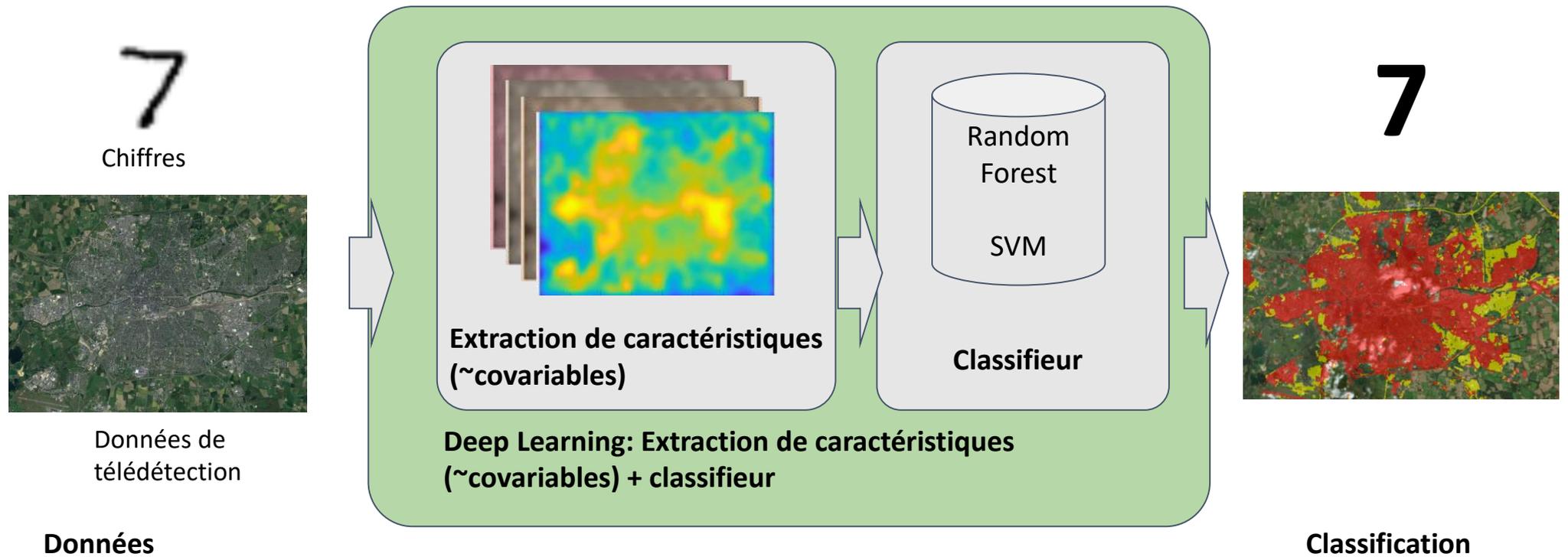
• Principe du Deep Learning (DL)

Les réseaux neuronaux en bref



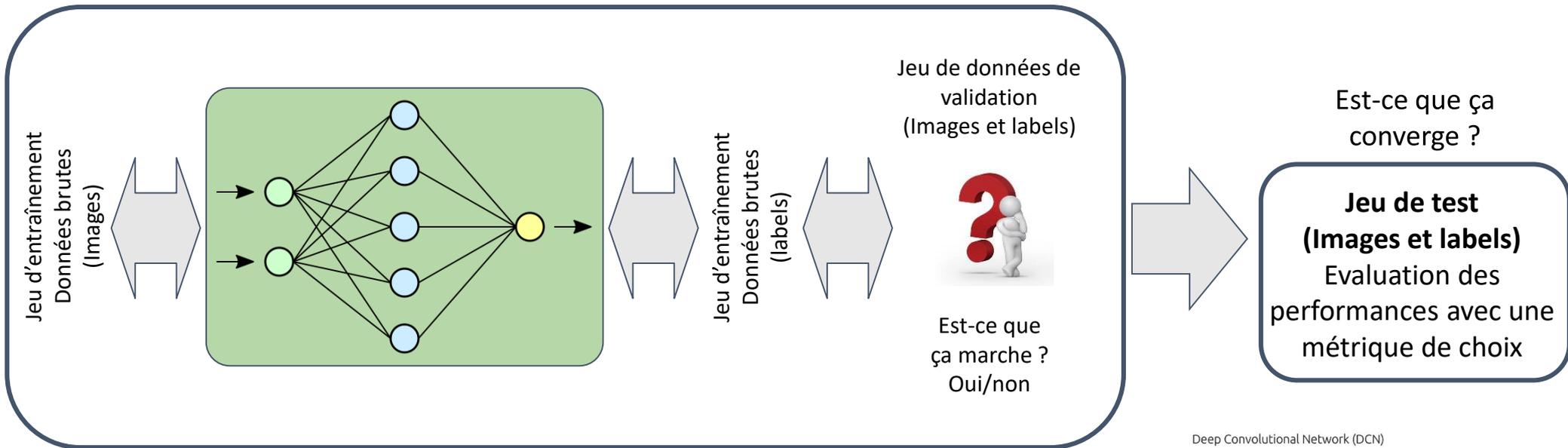
• Principe du Deep Learning (DL)

Les réseaux neuronaux en bref



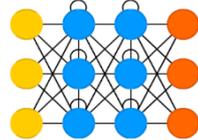
• Principe du Deep Learning (DL)

Les réseaux neuronaux en bref

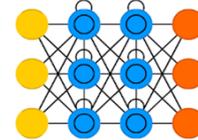


Plusieurs architectures de réseaux neuronaux pour différentes tâches

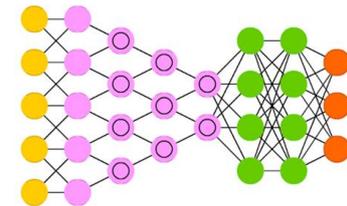
Recurrent Neural Network (RNN)



Long / Short Term Memory (LSTM)

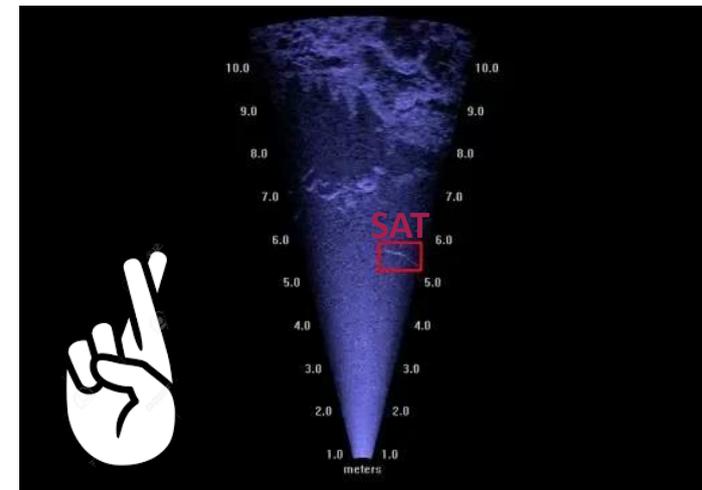


Deep Convolutional Network (DCN)



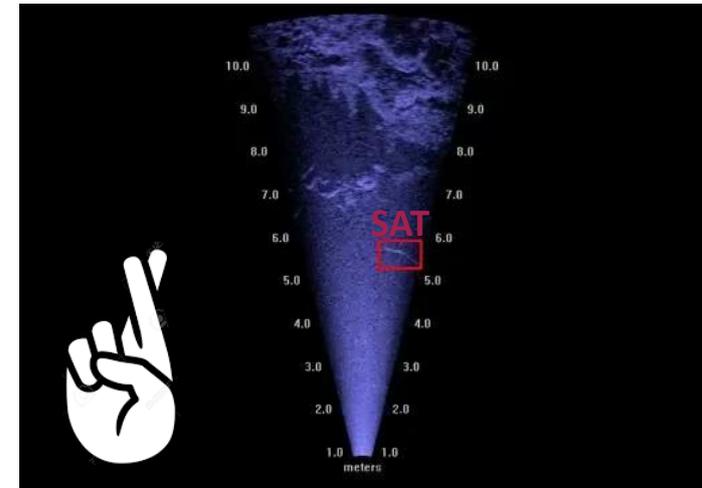
• Les différentes étapes du projet

- 1) Choix, développement et entraînement d'un réseau monoclasse pour détecter automatiquement le passage d'un poisson
- 2) Peut-on identifier automatiquement l'espèce du poisson ?
→ réseau multi-classe
- 3) Quelles sont les performances réelles de ces modèles ?
→ Jeu de données entièrement dépouillé (48h)
- 4) Diffuser ce modèle : développement d'une interface utilisateur open-source



• Les différentes étapes du projet

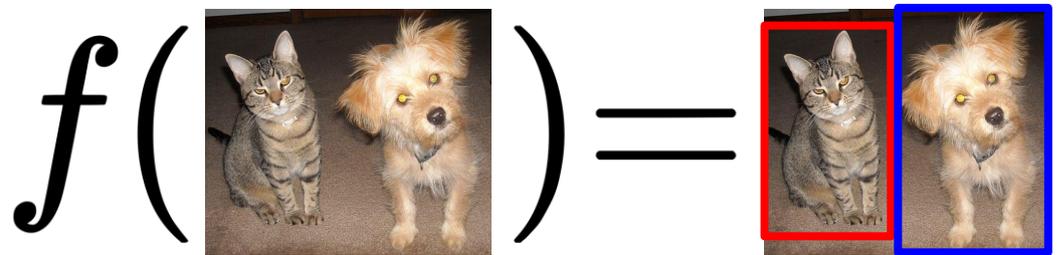
- 1) Choix, développement et entraînement d'un réseau monoclasse pour détecter automatiquement le passage d'un poisson
- 2) Peut-on identifier automatiquement l'espèce du poisson ?
→ réseau multi-classe
- 3) Quelles sont les performances réelles de ces modèles ?
→ Jeu de données entièrement dépouillé (48h)
- 4) Diffuser ce modèle : développement d'une interface utilisateur open-source



• Choix, développement et entraînement du réseau monoclasse

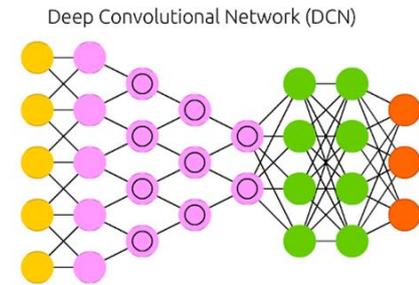
Choix du réseau neuronal : analyse d'image vs. analyse de vidéo

Chaque image est analysée séparément

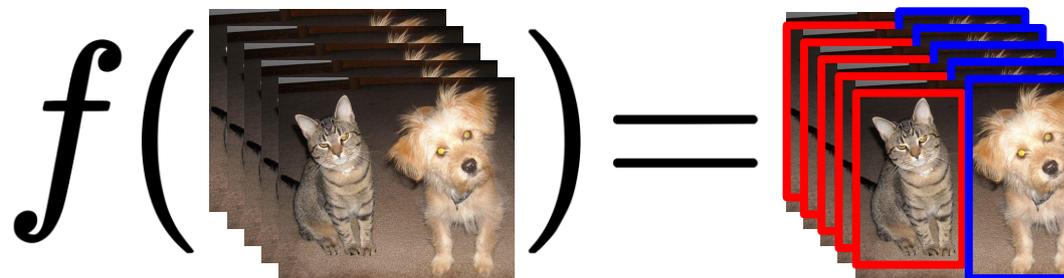


CNN: Convolutional NN

- 1 Bounding box par image
- 1 label par image
- Pas de corrélation temporelle

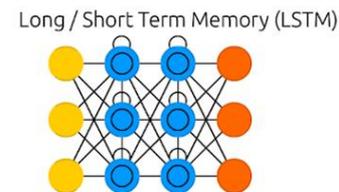


Tout le flux vidéo est analysé



RNN: Recurrent NN

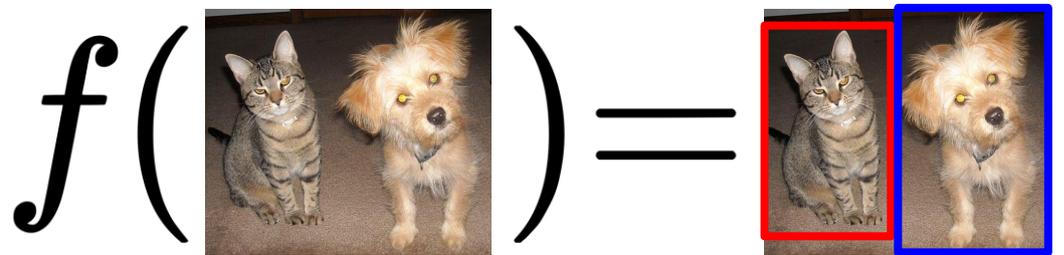
LSTM: Long/Shot term memory
+ CNN



• Choix, développement et entraînement du réseau monoclasse

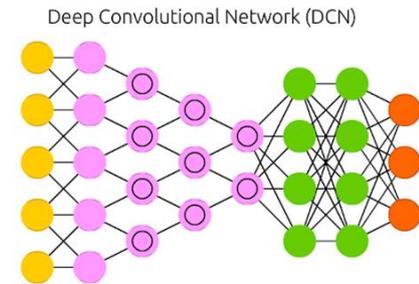
Choix du réseau neuronal : analyse d'image vs. analyse de vidéo

Chaque image est analysée séparément



CNN: Convolutional NN

- 1 Bounding box par image
- 1 label par image
- Pas de corrélation temporelle



Tout le flux vidéo est analysé

Pourquoi un CNN ?

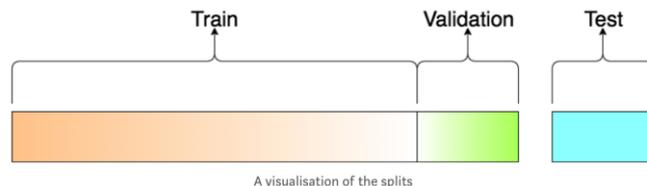
- Principe de simplicité : avant de développer des méthodes complexes, essayons les algorithmes déjà connus !
- Préparation des données : plus simple de produire de larges quantités d'images labellisées plutôt que de vidéos labellisées
- Analyse comparative : base de réseaux déjà entraînés, dans notre cas YOLO v3

• Choix, développement et entraînement du réseau monoclasse

Entraînement du CNN monoclasse

Jeu d'entraînement :

- 4 800 images labellisées (72 séquences vidéos) + 1 200 images vides
- Données ARIS (1/3) + données DIDSON (2/3)
- 1 seule classe = FISH
- Très forte variabilité (conditions environnementales, sites de suivi, tailles, espèces et comportements des poissons...)
- Entraînement (60%) + validation (18%) + test (22%)

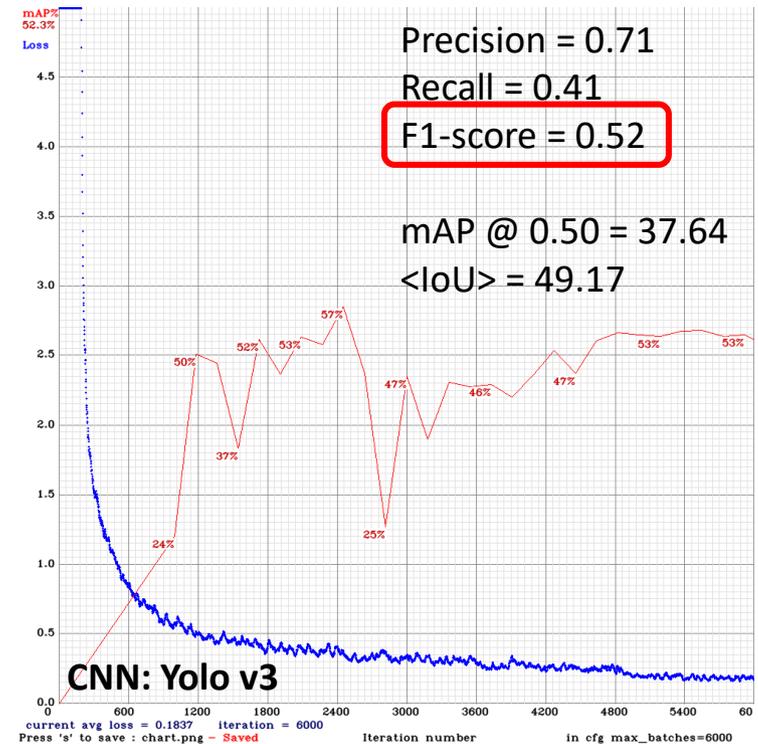
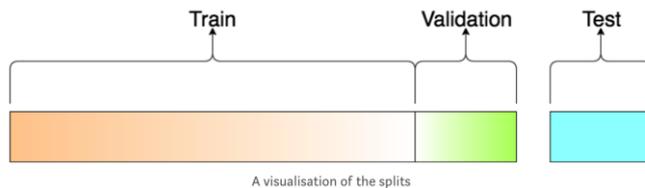


• Choix, développement et entraînement du réseau monoclasse

Entraînement du CNN monoclasse

Jeu d'entraînement :

- 4 800 images labellisées (72 séquences vidéos) + 1 200 images vides
- Données ARIS (1/3) + données DIDSON (2/3)
- 1 seule classe = FISH
- Très forte variabilité (conditions environnementales, sites de suivi, tailles, espèces et comportements des poissons...)
- Entraînement (60%) + validation (18%) + test (22%)



• Choix, développement et entraînement du réseau monoclasse

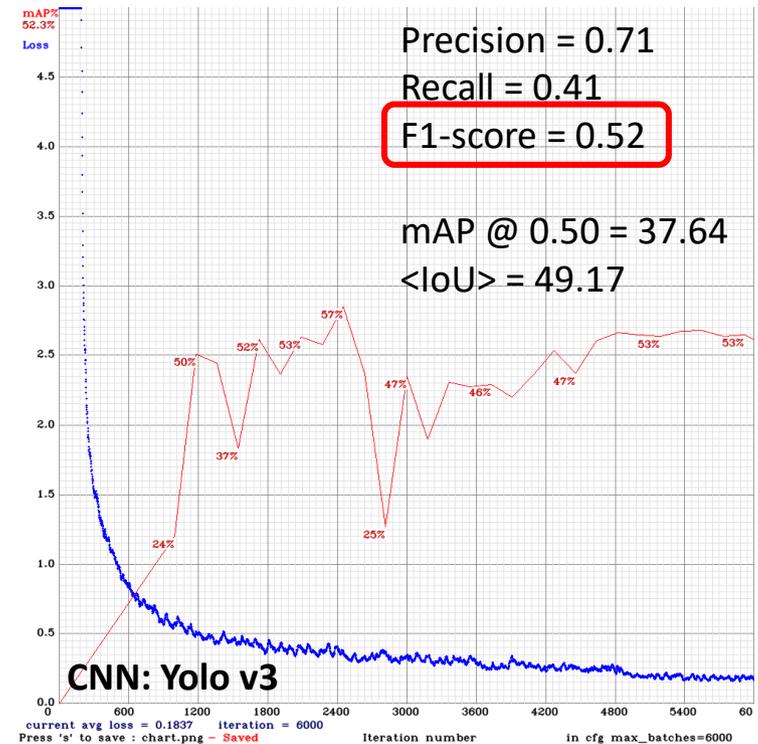
Entraînement du CNN monoclasse

Jeu d'entraînement :

- 4 800 images labellisées (72 séquences vidéos) + 1 200 images vides
- Données ARIS (1/3) + données DIDSON (2/3)
- 1 seule classe = FISH
- Très forte variabilité (conditions environnementales, sites de suivi, tailles, espèces et comportements des poissons...)
- Entraînement (60%) + validation (18%) + test (22%)

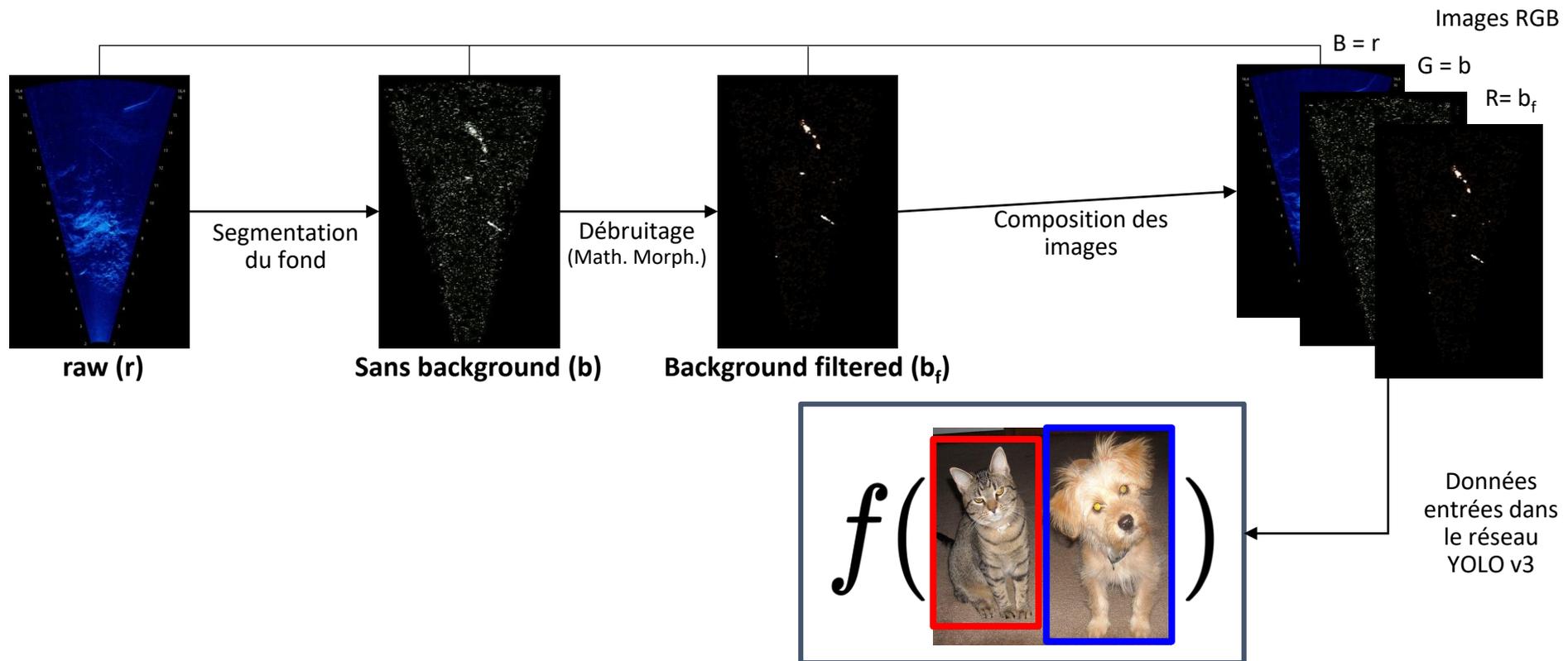


- Performances mitigées sur le jeu de test
- Données très bruitées
- Pré-traitement des images pour favoriser la détection des poissons par l'algorithme



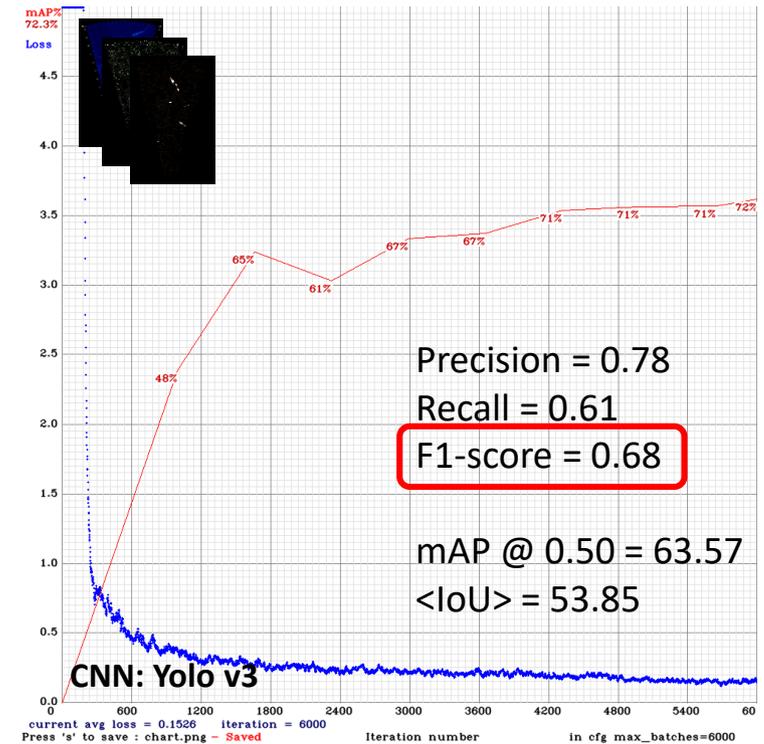
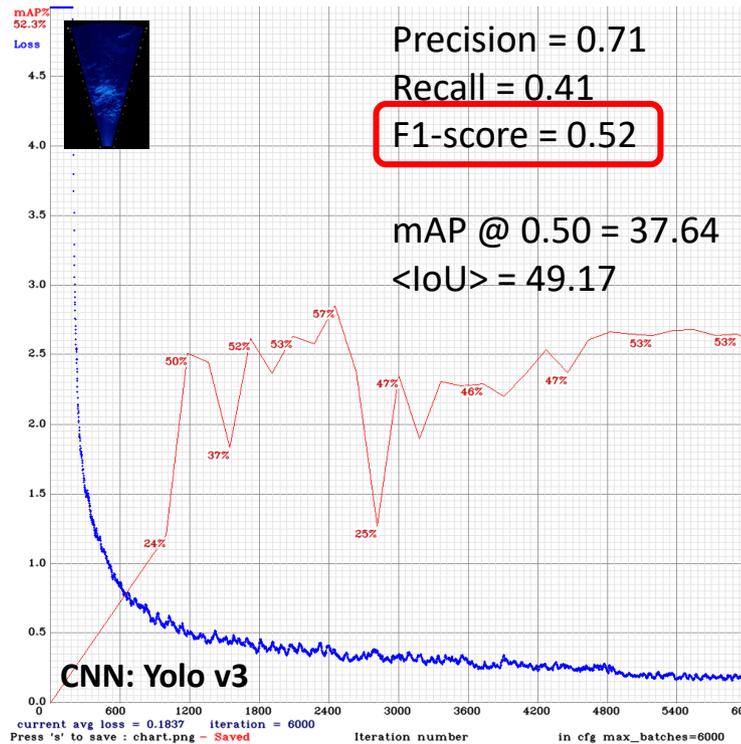
• Choix, développement et entraînement du réseau monoclasse

Développement de la pipeline AcousticAI (des vidéos brutes aux détections)



• Choix, développement et entraînement du réseau monoclasse

Développement de la pipeline AcousticAI (des vidéos brutes aux détections)



• Choix, développement et entraînement du réseau monoclasse

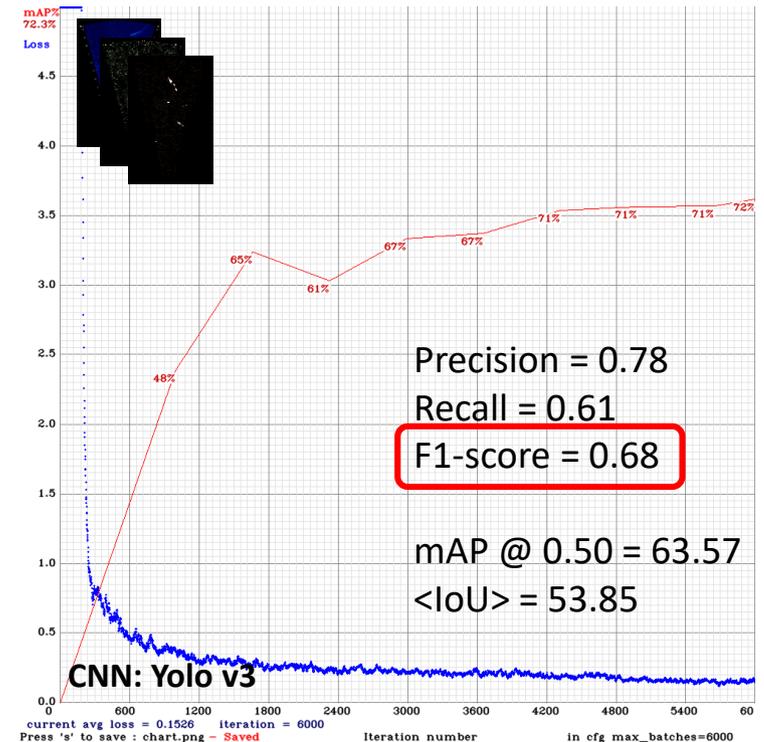
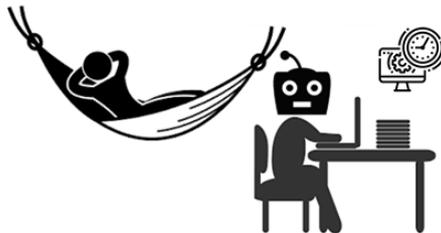
Développement de la pipeline AcousticAI (des vidéos brutes aux détections)

		Vérité terrain	
		Poisson	Autre
CNN monoclasse	Poisson	0.61	0.13
	Autre	0.39	0.87

Normalisé par colonne

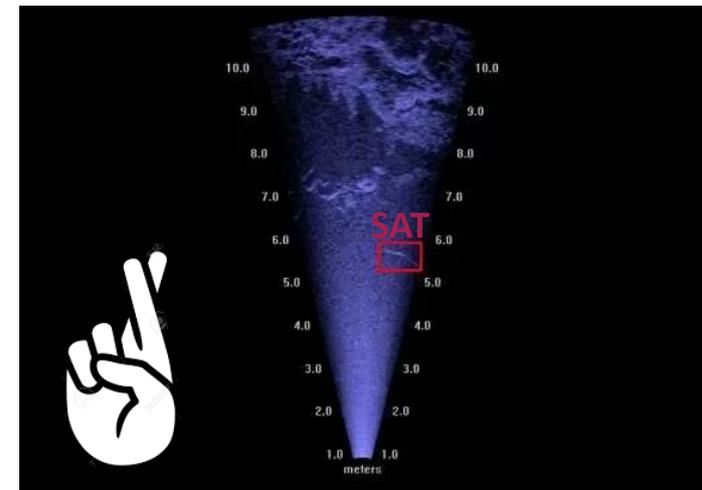
(donne l'erreur par rapport à la vérité terrain)

- 61% des poissons du jeu de test sont bien détectés
- 87% des images vides peuvent être éliminés



• Les différentes étapes du projet

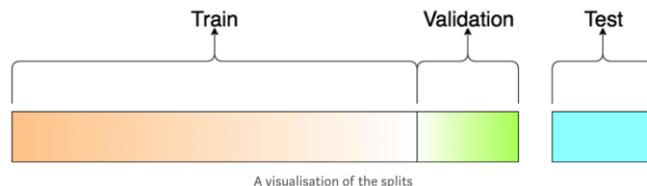
- 1) Choix, développement et entraînement d'un réseau monoclasse pour détecter automatiquement le passage d'un poisson
→ réseau multi-classe
- 2) Peut-on identifier automatiquement l'espèce du poisson ?
→ réseau multi-classe
- 3) Quelles sont les performances réelles de ces modèles ?
→ Jeu de données entièrement dépouillé (48h)
- 4) Diffuser ce modèle : développement d'une interface utilisateur open-source



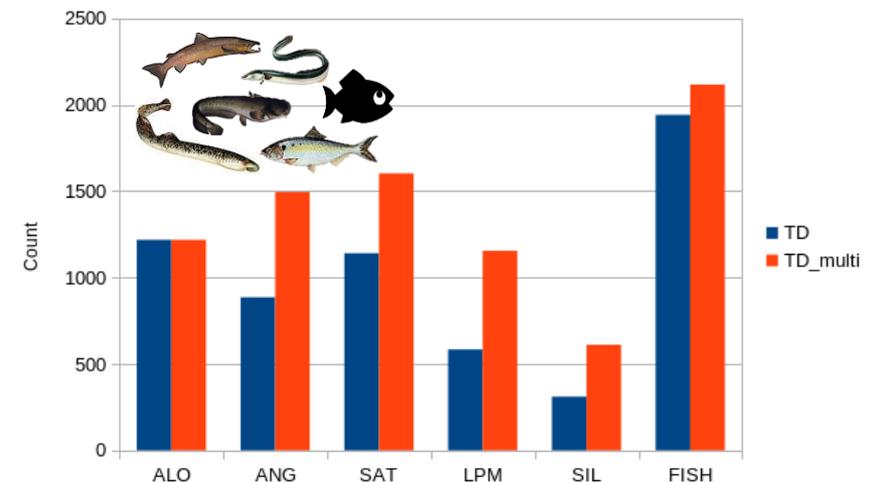
• Identifier automatiquement l'espèce du poisson : le réseau multi-classe

Jeu d'entraînement :

- **4 800** images labellisées (72 séquences vidéos) + 1 200 images vides
- Données ARIS (1/3) + données DIDSON (2/3)
- **1 seule classe = FISH**
- Très forte variabilité (conditions environnementales, sites de suivi, tailles, espèces et comportements des poissons...)
- Entraînement (60%) + validation (18%) + test (22%)



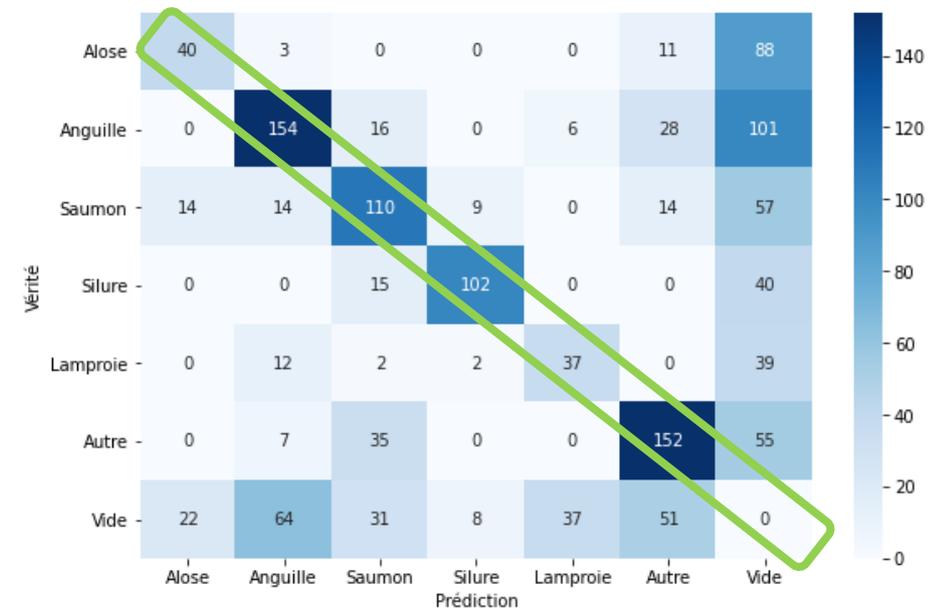
- **9 000 images**
- **ARIS (1/2) + DIDSON (1/2)**
- **6 classes = espèces d'intérêt**



• Identifier automatiquement l'espèce du poisson : le réseau multi-classe

Performances sur le jeu de test

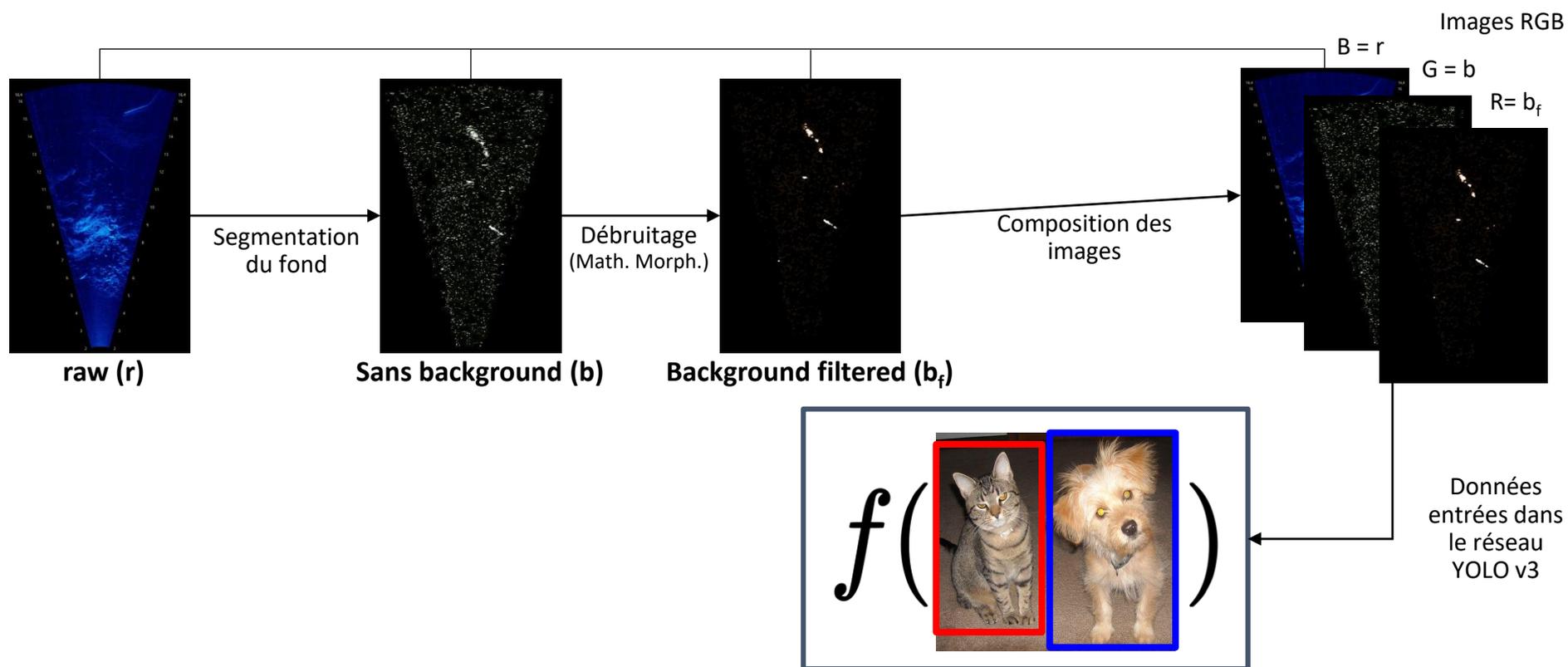
Classe	Precision	Recall	mAP@.5	F1-score
Silures (SIL)	0.84	0.65	0.63	0.73
Autres (FISH)	0.60	0.61	0.49	0.60
Anguilles (ANG)	0.62	0.51	0.45	0.56
Saumons (SAT)	0.57	0.54	0.33	0.55
Lamproies marines (LPM)	0.48	0.40	0.25	0.44
Aloses (ALO)	0.53	0.28	0.18	0.37
Moyenne	0.61	0.50	0.39	0.55



- Métriques plus faibles pour la classification que pour la détection,
- D'après la matrice de confusion = peu de confusion entre classes, mais problème de FP et de FN,
- Performances sur des images fixes = pas de cohérence temporelle = post-traitement ?

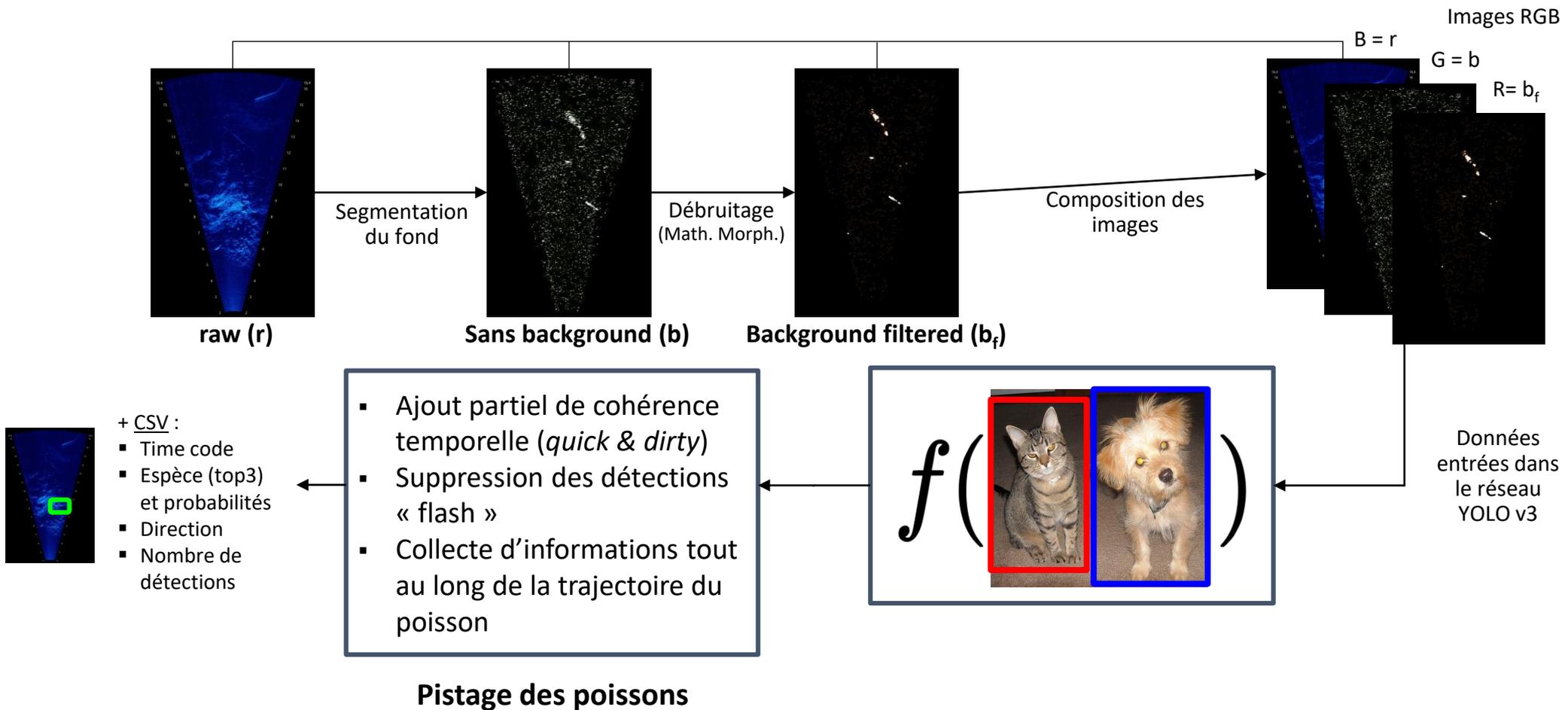
• Identifier automatiquement l'espèce du poisson : le réseau multi-classe

Développement de la pipeline AcousticAI (des vidéos brutes aux détections)



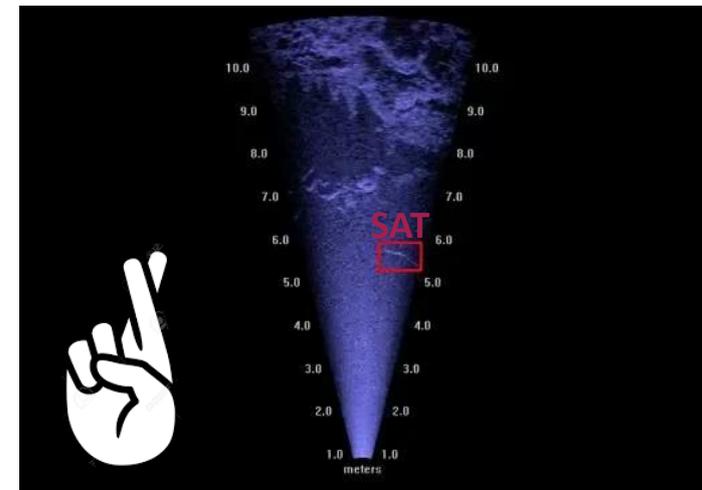
• Identifier automatiquement l'espèce du poisson : le réseau multi-classe

Développement de la pipeline AcousticAI (des vidéos brutes aux détections)



• Les différentes étapes du projet

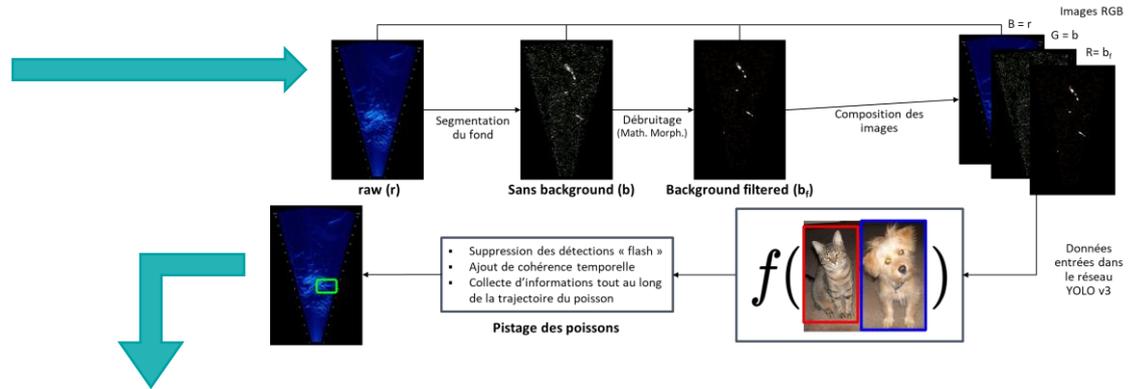
- 1) Choix, développement et entraînement d'un réseau monoclasse pour détecter automatiquement le passage d'un poisson
- 2) Peut-on identifier automatiquement l'espèce du poisson ?
→ réseau multi-classe
- 3) Quelles sont les performances réelles de ces modèles ?
→ Jeu de données entièrement dépouillé (48h)
- 4) Diffuser ce modèle : développement d'une interface utilisateur open-source



• Performances sur le Validation Dataset

Jeu de données de validation (VD)

- 80 vidéos ARIS + 104 vidéos DIDSON = 40h
- vidéos vides : 27 DIDSON + 6 ARIS = 8h,
- lecture intégrale des fichiers par un opérateur
- description de chaque passage de poisson
- 1 856 poissons = 1 047 DIDSON + 809 ARIS



csv
+ mp4

Espèce	caméra	Taille du poisson				Total
		20-40	40-60	60-80	>80	
Aloses	Didson	0	1	0	0	1
	Aris	0	0	0	0	0
Anguilles	Didson	0	14	15	3	32
	Aris	13	52	15	2	82
Saumons	Didson	0	14	33	6	53
	Aris	0	2	10	0	12
Silures	Didson	0	0	0	2	2
	Aris	0	0	0	6	6
Lamproies	Didson	0	0	2	2	4
	Aris	0	0	0	1	1
Poissons génériques	Didson	315	505	30	2	852
	Aris	248	419	18	0	685

• Performances sur le Validation Dataset

	Taille du poisson				Total
	20-40	40-60	60-80	>80	
Yolov5 Multiclasse	29.27	55.53	76.25	86.67	45.70
Yolov3 Monoclasse	57.62	76.22	90.00	93.33	67.91

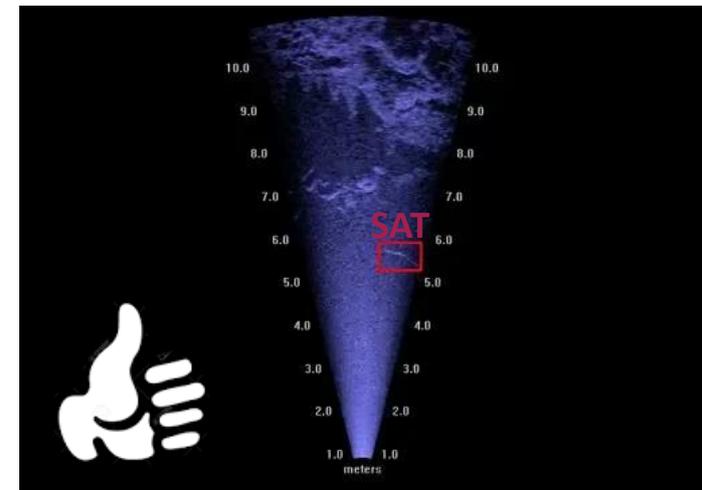
- Plus les poissons sont gros, mieux ils sont détectés
- Réseau mono-classe détecte plus de poissons que le multi-classe

		Saumons (n= 65)	Anguilles (n = 114)
Détecté	Total	77.4% (n=50)	75.0 (n=85)
	Pas détecté	22.6% (n=15)	25.0% (n=29)

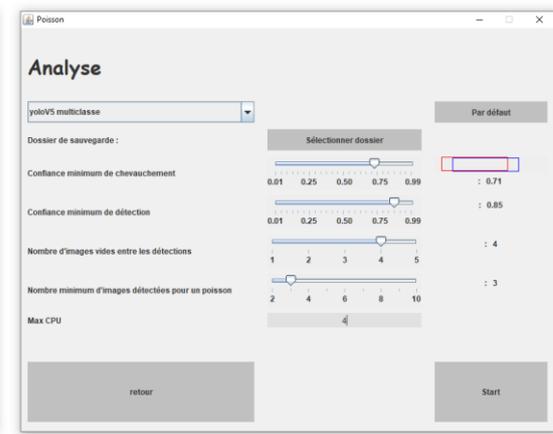
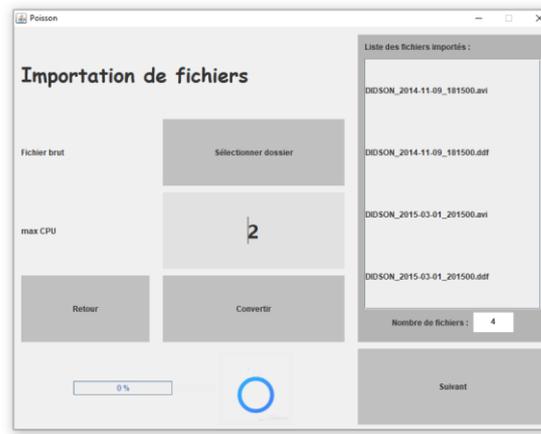
- Peu d'aloses, silures et lamproies marines dans le VD
- Saumons et anguilles = bon taux de détection (75%) et résultats prometteurs sur la classification !
... mais taux de faux positifs très important :
 - 1 089 FP saumons (precision = 0.03)
 - 1 227 FP anguilles (precision = 0.04)
- Moins chronophage qu'une lecture intégrale des vidéos pour un opérateur, mais pas encore optimal.

• Les différentes étapes du projet

- 1) Choix, développement et entraînement d'un réseau monoclasse pour détecter automatiquement le passage d'un poisson
→ réseau multi-classe
- 2) Peut-on identifier automatiquement l'espèce du poisson ?
→ réseau multi-classe
- 3) Quelles sont les performances réelles de ces modèles ?
→ Jeu de données entièrement dépouillé (48h)
- 4) Diffuser ce modèle : développement d'une interface utilisateur open-source



• Développement d'une interface utilisateur open-source



- Compatible Windows/Linux, français et anglais
- Vidéos brutes → vidéos converties → pipeline → CSV et vidéos traitées
- Parallélisation des pré-traitements (CPU), puis analyse sur GPU (si disponible)
- Possibilité de choisir le réseau de neurones et d'en intégrer de nouveaux,
- Post-traitement paramétrable par l'opérateur (nombre de détection de chaque poisson, confiance minimale...)
- Plus de fonctionnalités à venir : création de clips...

Valid. Dataset (48h)	
Sans GPU	Avec GPU
48h	6h



• Conclusion générale



Automatiser la détection des poissons pour éviter une lecture intégrale des données par les opérateurs

➤ Seulement 46% des poissons sont bien détectés avec le réseau multi-classe (68% avec le mono-classe)

MAIS



Manquer un minimum de poissons d'espèce d'intérêt = minimiser les taux de FN

➤ 77% des saumons et 75% des anguilles sont bien détectés



Développer un outil cross-caméras acoustiques

➤ Résultats DIDSON > résultats ARIS (*pas présenté ici*)



Minimiser le temps opérateur = maximiser le taux de TN

➤ Une large majorité des images vides est correctement éliminée



Identifier automatiquement les poissons d'espèces d'intérêt

➤ Encourageant pour les saumons (70%) et les anguilles (77%) : analyse à poursuivre pour mieux comprendre les performances (tailles, position...)



Rendre possible l'utilisation du réseau par la communauté des utilisateurs de caméras acoustiques

➤ Développement d'une interface open-source

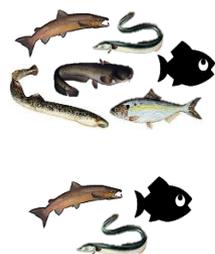
- **Premier réseau de neurones cross-camera et multi-espèces (Fernandez Garcia et al., in prep.)**
- **Modèle prêt à être utilisé par les opérateurs, sans nécessité de savoir coder ...**

... mais performances à améliorer en terme de détection et d'identification

• Perspectives : améliorer les performances de la pipeline

Améliorer le réseau de neurones :

- Inclure explicitement l'aspect temporel dans le CNN : détection d'action, réseaux récurrents (LSTM-CNN : ROLO?)
- Adaptation de domaine pour les données des caméras acoustiques ?
- Réduire le nombre de classes du réseau multi-classe : stage d'Antoine Lebeaud (2022)



	ARIS		DIDSON	
	Top 3	Top 1	Top 3	Top 1
précision	0.03	0.05	0.03	0.02
recall	0.33	0.33	0.55	0.28
F1 score	0.06	0.09	0.06	0.04
précision	0.06	0.04	0.1	0.06
recall	0.25	0.25	0.25	0.21
F1 score	0.1	0.07	0.14	0.09

Résultats préliminaires sur les 65 saumons du VD

- Augmentation du F1-score
- Nombre de FP divisé par 4 mais nombre de FN supérieur

Améliorer le tracking :

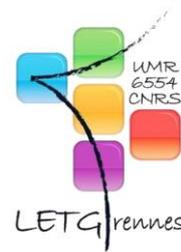
- Optimiser le paramétrage : nombre de détections consécutives, confiance minimale du réseau
- Intégrer l'expertise biologique : phénologie, rythme d'activité...
- Collecter plus d'informations relatives aux populations : mesure de la taille des individus (segmentation sémantique)

Diffuser et déployer le réseau aux utilisateurs des caméras acoustiques pour favoriser l'augmentation du niveau de connaissances des populations de diadromes et leur gestion

Merci de votre attention,

Et merci à :

Laurent Beaulaton
Laureline Blivet
Thomas Corpetti
Olivier Dezerald
Marc Hersant
Sami Jabbalah
Antoine Lebeaud
Marie Nevoux



Contacts : francois.martignac@inrae.fr
guglielmo.fernandez-garcia@univ-rennes2.fr