



Cet article est publié sous la licence Creative Commons (CC BY 4.0). La citation comme l'utilisation de tout ou partie du contenu de cet article doit obligatoirement mentionner les auteurs, l'année de publication, le titre, le nom de la revue, le volume, le numéro de l'article et le DOI.

## Automatisation de l'analyse de données des caméras acoustiques par un réseau de neurones convolutifs

François MARTIGNAC<sup>1,4</sup>, Guglielmo FERNANDEZ-GARCIA<sup>1,2,4</sup>, Thomas CORPETTI<sup>2</sup>, Marie NEVOUX<sup>1,4</sup>, Laurent BEAULTON<sup>3,4</sup>

<sup>1</sup> DECOD (Dynamique et Durabilité des Ecosystèmes), L'Institut Agro, IFREMER, INRAE, Rennes, France.

<sup>2</sup> UMR LETG, CNRS-Université Rennes 2, Rennes, France.

<sup>3</sup> Service Conservation et gestion durable des espèces exploitées, Direction de la Recherche et de l'Appui Scientifique, OFB, U3E, Rennes, France.

<sup>4</sup> Pôle Gestion des Migrateurs Amphihalins dans leur Environnement, OFB, INRAE, L'Institut Agro, UPPA, Rennes, France.

Correspondance : François MARTIGNAC, [francois.martignac@inrae.fr](mailto:francois.martignac@inrae.fr)

*Les caméras acoustiques rendent possible l'accès à de nouvelles informations sur la biologie, la morphologie et le comportement de poissons migrateurs en rivière, tout en restant non intrusives. Elles enregistrent d'importants jeux de données dont l'analyse très chronophage pourrait être optimisée par des méthodes d'intelligence artificielle, telle que celle présentée dans cette étude. Le réseau de neurones convolutifs permet une détection automatique des poissons et se révèle prometteur pour identifier leur espèce.*

### Contexte et objectifs

Les méthodes hydroacoustiques sont des méthodes novatrices en plein développement à l'échelle nationale et internationale dans le cadre d'études environnementales et de suivi des populations de poissons. En effet, les avancées technologiques permettent d'augmenter la couverture du volume d'eau échantillonné, de maximiser la résolution des images ou encore de disposer d'outils de moins en moins volumineux, favorisant leur utilisation dans un maximum d'applications différentes, sans interférer ni avec le comportement, ni avec l'intégrité physique des poissons. L'avènement des caméras acoustiques, mises au point il y a une vingtaine d'années, illustre bien cette évolution : grâce aux très hautes fréquences utilisées et à l'augmentation du nombre de faisceaux, les données acoustiques enregistrées par ces outils sont projetées sous forme d'images dont la résolution est proche de celle d'une vidéo subaquatique. Ainsi, la morphologie et le comportement du poisson sont facilement observables sur les données, ce qui oriente l'identification de son espèce (Martignac *et al.*, 2015).

Cette technologie intéresse donc un public de plus en plus large à l'échelle internationale : la communauté scientifique, les gestionnaires des milieux aquatiques, les prestataires privés, tout comme les grandes sociétés privées qui doivent respecter les réglementations

relatives aux problématiques de continuité écologique. Cependant, malgré leurs nets avantages, les principales limites de ces méthodes restent le volume important des données enregistrées et, surtout, l'analyse de ces vidéos, très chronophage en l'absence d'outils automatiques de lecture des vidéos acoustiques. Différents outils ont été développés ces dernières années, mais aucune ne s'attache à simultanément détecter les passages de poissons et identifier plusieurs espèces (Wei *et al.*, 2022). Cette étude vise à développer une nouvelle méthode automatisée d'analyse d'images provenant de différentes caméras acoustiques par un réseau de neurones profonds. L'objectif est de détecter et classer les poissons par espèce, en créant un outil accessible à tous les utilisateurs de caméras acoustiques. Les caméras acoustiques DIDSON et ARIS, toutes deux développées par Sound Metrics Corp. (États-Unis) sont les deux modèles les plus utilisés à l'échelle internationale (Martignac *et al.*, 2015). Issues de deux générations différentes d'un même outil, les similarités des vidéos qu'elles enregistrent motivent le souhait de développer un outil d'analyse commun pour ces deux modèles. À l'avenir, l'outil pourrait être transposé à des caméras acoustiques qui produisent des images de taille et de résolution différentes (notamment, les caméra Oculus, Blueprint Subsea et BlueView, Teledyne Inc.), décrites dans l'encadré 1.

## Jeux de données

### Les réseaux de neurones profonds

Les algorithmes d'apprentissage profond (en anglais : « *Deep Learning* ») apprennent à reconnaître des patrons au sein de la donnée afin de la classifier (LeCun *et al.*, 2015). Ils ont prouvé leur utilité dans de nombreux domaines, notamment en écologie (Christin *et al.*, 2019 ; Lamba *et al.*, 2019). En effet, ils permettent de réduire le coût des calculs et de surmonter les problèmes numériques et conceptuels liés à la complexité des données et des phénomènes à analyser. Ils ont aussi des limites qui méritent d'être soulignées : ce sont des méthodes empiriques qui peuvent impliquer une perte de compréhension des phénomènes sous-jacents, et qui dépendent fortement de la qualité des données fournies au cours de la phase d'apprentissage (O'Neil, 2016). Il existe un nombre croissant de types de réseaux neuronaux, selon la structure différente des « neurones » et des « couches profondes » : les paramètres des « neurones » indivi-

duels sont en fait transmis à la « couche » suivante et le schéma de connexion définit l'architecture du réseau. On les appelle « réseaux convolutionnels » (CNN, *Convolutional Neural Network*), ces architectures dans lesquelles tous les paramètres des « neurones » d'une « couche » donnée sont transmis à tous les « neurones » de la « couche » suivante. Ce type de réseaux est particulièrement adapté à la détection d'objet, tant pour les images que pour la vidéo. Le réseau de neurones utilisé dans cette étude est un CNN appelé YOLO (Jocher, 2020), développé pour la reconnaissance d'objets au sein d'images.

### Jeux de données d'entraînement

Afin d'entraîner un CNN, il est nécessaire de disposer d'un grand nombre d'images dans lesquelles des experts ont déjà identifié la présence (ou non) d'un objet d'intérêt et sa position. Ces jeux de données labellisés sont constitués d'un certain nombre d'images qui peuvent aller de quelques centaines à des milliers. En outre, pour

#### Encadré 1 – Des modèles de caméras acoustiques adaptés à diverses applications.

Plusieurs modèles de caméras acoustiques sont actuellement disponibles et cette technologie tend à se démocratiser. Les quatre modèles de caméras acoustiques les plus couramment utilisées, regroupées en deux familles, seront présentées ici : celles qui favorisent la finesse de la résolution des images enregistrées (ou taille des pixels) et celles qui maximisent le volume échantillonné. Les caméras acoustiques les plus anciennes, les DIDSON et ARIS (Sound Metrics Corp., États-Unis), appartiennent à la première famille tandis que les BlueView (Teledyne Inc., CA, États-Unis) et Oculus (Blueprint Subsea, Royaume-Uni) à la seconde, grâce à une ouverture de leur champ de vision de 130° (contre 29° pour les DIDSON et ARIS). Chacune de ces caméras possède un mode basse-fréquence, qui maximise sa portée mais minimise sa résolution, et un mode haute fréquence qui maximise la résolution des images, sur une portée plus limitée. Le nombre de faisceaux, impactant la résolution, diffère également entre les modèles (tableau 1).

**Tableau 1** – Caractéristiques des quatre modèles de caméras acoustiques les plus couramment utilisés dans le cadre de suivis environnementaux (LF : basse fréquence ; HF : haute fréquence).

	DIDSON Std	ARIS 3000	BlueView M900-2250-130-Mk2	Oculus M750d
Constructeur	Sound Metrics C.	Sound Metrics C.	Teledyne Marine	Bleprint Subsea
Fréquences LF/HF (MHz)	1,10/1,80	1,80/3,00	0,90/2,25	0,75/1,20
Fenêtre max (m)	40/10	15/5	100/10	120/40
Nombre de faisceaux	48/96	128	768	512
Ouverture horizontale	29°	29°	130°	130°
Ouverture verticale	14°	14°	12°/20°	20°
Nombre maximal d'images par seconde	21	15	25	40
Dimensions (cm)	31 x 21 x 17	26 x 16 x 14	21 x 13 x 10	13 x 12 x 6
Poids (kg)	7,90	5,12	2,50	0,98
Coût	+++	+++	++	++

Par leurs différences de caractéristiques intrinsèques, l'utilisation de ces caméras peut différer. La résolution supérieure des DIDSON et ARIS favorise l'identification des espèces sur des critères morphologiques et le suivi d'individus de taille réduite à moyenne sur des fenêtres d'acquisitions réduites. À l'inverse, la détection et le suivi d'individus de taille importante (>50 cm) sont facilités sur les caméras BlueView et Oculus, grâce à des volumes couverts bien plus importants. De plus, les caméras BlueView et Oculus permettent d'enregistrer un nombre d'images à la seconde plus important que les caméras DIDSON et ARIS, ce qui rend possible leur utilisation en mobile, à vitesse réduite. La portée, le nombre de faisceaux et le nombre d'images par seconde impactent le volume des fichiers enregistrés : les vidéos enregistrées par les caméras acoustiques BlueView et Oculus sont encore plus lourdes que celles enregistrées par les DIDSON et ARIS. Les logiciels de post-traitement des vidéos Sound Metrics sont nettement plus riches que ceux des deux autres caméras, qui ne sont que des lecteurs basiques. Enfin, les dimensions plus réduites des BlueView et Oculus, rendent leur maintenance plus facile par les opérateurs, et leur coût est nettement inférieur aux caméras Sound Metrics. Au regard de tous ces éléments, les caméras BlueView et Oculus sont à privilégier pour la détection et les suivis en temps réel d'individus de grande taille, alors que les caméras DIDSON et ARIS exploitent plus largement leurs avantages en utilisation fixe sur des acquisitions continues long-terme.

que les performances du modèle ne soient pas dépendantes des données d'apprentissage, il est important que les images choisies reflètent la variabilité des images auxquelles on souhaite appliquer le modèle. Cette diversité inclut ici, le type de caméra utilisée (DIDSON ou ARIS) avec leur disparité de résolution (figure 1), les différentes classes d'objet présents dans les données disponibles ainsi que leurs variations de morphologie, de comportement de nage, les distances entre le poisson et la caméra, les conditions environnementales lors de l'enregistrement (débits, turbidités). Pour assurer une bonne transposition du réseau de neurones à différents sites de suivi, les enregistrements utilisés proviennent de deux installations différentes. Des images sans aucun passage de poisson figurent également dans ce jeu de données.

Le jeu de données d'entraînement (tableau 2) se compose de plus de 6 800 images contenant des passages de poissons (3 280 images DIDSON, extraites de 41 vidéos; 3 540 images ARIS, extraites de 102 vidéos), complétées de 6 310 images vides (3 080 DIDSON et 3 230 ARIS). Les espèces sélectionnées sont celles qui vont être majoritairement recherchées par les utilisateurs français du réseau sur les données enregistrées par les caméras acoustiques, à savoir les stades adultes des saumons atlantiques (*Salmo salar*), anguilles européennes (*Anguilla anguilla*), lamproies marines (*Petromyzon marinus*) et silures glanes (*Silurus glanis*). Les espèces, qui sont difficiles à identifier de manière formelle, ou qui présentent un intérêt moindre en termes de gestion ou de connaissances scientifiques ont été regroupées dans une classe « autres espèces ». Le jeu de données est divisé en trois groupes distincts : i) un jeu d'entraînement, permettant l'optimisation du réseau de neurones à la détection de poissons (60% environ du jeu de données total); ii) un jeu de validation ayant pour but d'évaluer l'erreur du modèle pendant l'entraînement (20% du jeu de données total); iii) un jeu de test, permettant d'évaluer les performances du modèle à la fin de l'entraînement (20% du jeu de données total).

### Évaluation des performances du réseau

L'efficacité du réseau en conditions réelles est évaluée en le déployant sur un jeu de données de validation. Celui-ci est composé de cent-quatre vidéos de quinze minutes de la caméra DIDSON et de quatre-vingt vidéos de dix minutes de la caméra ARIS pour un total de près de quarante heures de vidéos à cinq images par seconde. Dans ce jeu de données également, les vidéos sont variées, provenant de deux sites différents et de différentes périodes de l'année, avec sur chaque séquence un nombre variable de poissons, d'objets dérivants, de plantes aquatiques fixées sur le lit du cours d'eau. Un jeu de huit heures de vidéos sans aucun passage de poisson est ajouté à ce jeu de validation. Toutes ces vidéos ont été entièrement visionnées par un opérateur expérimenté qui a annoté chaque passage de poisson, identifié l'espèce lorsque celle-ci pouvait l'être avec certitude et évalué la classe de taille de l'individu (tableau 3). On considère que l'expert est efficace à 100% dans la détection comme dans l'identification des espèces de poissons. Aucune séquence de ces vidéos n'a été utilisée au cours de la phase d'apprentissage du modèle.

Les passages de saumons et d'anguilles, espèces à fort intérêt, sont volontairement beaucoup plus représentés

Figure 1 – Images extraites de vidéos enregistrées par les caméras acoustiques ARIS (à gauche) et DIDSON (à droite). Les cadres rouges identifient les échos renvoyés par deux poissons.

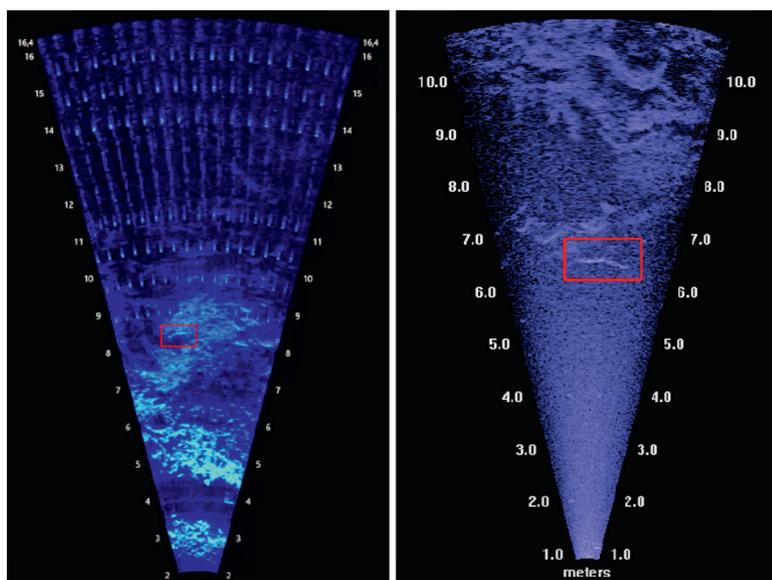


Tableau 2 – Répartition des images labellisées et du nombre de séquences vidéos en fonction des espèces identifiées et des modèles de caméras acoustiques dont elles sont extraites.

Classes	Interne		Externe	
	ARIS	DIDSON	ARIS	DIDSON
Anguilles européennes	845	583	26	5
Saumons atlantiques	648	846	27	11
Silures glanes	971	151	20	4
Lamproies marines	300	284	21	7
Autres espèces	772	1 420	8	14
Vides	3 232	3 077	38	10
Total	6 768	6 361	140	51

Tableau 3 – Répartition des objets détectés par l'opérateur entre les différentes classes dans le jeu de validation en fonction du type de caméra et de leur classe de taille (en cm).

Classes	Caméra	Taille du poisson (cm)				Total	
		20-40	40-60	60-80	> 80		
Anguilles européennes	DIDSON	0	14	15	3	32	112
	ARIS	13	52	15	0	80	
Saumons atlantiques	DIDSON	0	14	33	6	53	65
	ARIS	0	2	10	0	12	
Silures glanes	DIDSON	0	0	0	2	2	8
	ARIS	0	0	0	6	6	
Lamproies marines	DIDSON	0	0	0	2	4	5
	ARIS	0	0	0	1	1	
Autres espèces	DIDSON	315	506	30	2	853	1538
	ARIS	248	419	18	0	685	
Total	DIDSON	315	534	80	15	944	1728
	ARIS	261	473	43	7	784	

que ceux des silures et des lamproies dans le jeu de validation. Au total, les passages de 1 728 poissons ont été identifiés et décrits par l'opérateur (944 sur les fichiers DIDSON et 784 sur les fichiers ARIS, tableau 3).

### Pipeline d'analyse des données

L'objectif de l'étude est de détecter les signaux renvoyés par des échos de poissons parmi le fort bruit présent dans les images enregistrées par les caméras acoustiques, et de pouvoir suivre l'intégralité du passage du poisson dans le faisceau. Ainsi, au-delà de l'analyse des données par le réseau de neurones, la méthode proposée lors de ce travail est de prétraiter les données grâce à de la vision par ordinateur classique pour aider le réseau de neurones à différencier le poisson du bruit de fond, puis d'introduire de la cohérence temporelle entre les détectations image par image appartenant à une même séquence vidéo. L'intégralité du pipeline<sup>1</sup> est présentée en figure 2.

#### Prétraitement des données

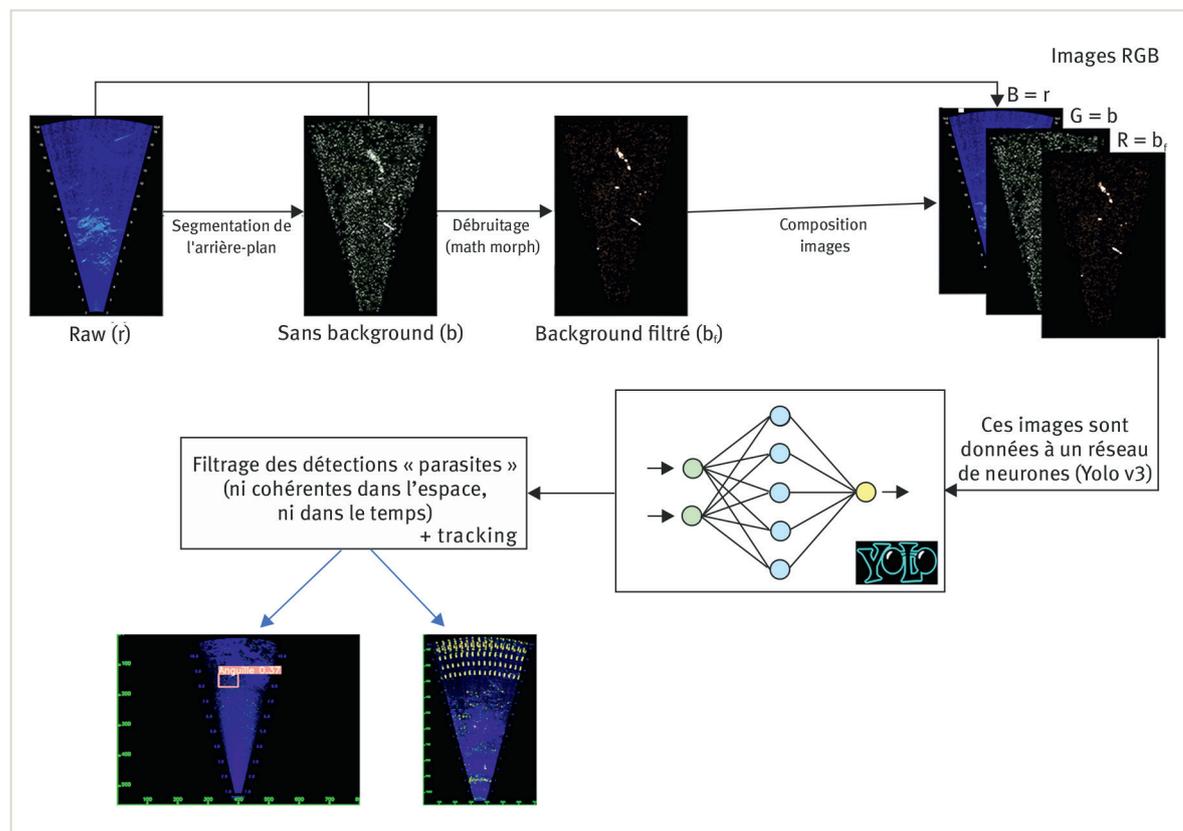
Suite à la conversion des données acoustiques au format vidéo, lors du prétraitement, celles-ci sont segmentées en images successives: le réseau de neurones analyse seulement les images des vidéos, et non les vidéos en elles-mêmes. Une réduction du bruit au sein de la donnée est nécessaire car les images contiennent beaucoup d'informations non désirables. En effet, le fond du cours d'eau émet un écho fixe avec une intensité plus importante

que celle d'un poisson, et peut donc gêner sa détection. Le prétraitement des images est réalisé en deux étapes: utilisation d'un mélange gaussien (Zivkovic et Van Der Heijden, 2006) puis un dé-bruitage utilisant un filtre médian (Ahmad et Sundararajan, 1987) et opérateur d'ouverture (Dougherty et Lotufo, 2003). Cela permet de garder uniquement les objets en mouvements sur notre image, et de masquer les informations non nécessaires pour le modèle. En recomposant une nouvelle image à partir de l'image initiale et des deux images créées par le prétraitement, on obtient une image trois canaux qui peut être fournie au réseau, favorisant son apprentissage et ses performances de détection.

#### Analyse des données par le réseau de neurones

Les trois images (ou image trois canaux) sont donc analysées par le réseau de neurones. En sortie, le modèle crée des labels de prédiction et des coordonnées de boîte englobante<sup>2</sup> pour chacune des détectations. Étant donné que le réseau de neurones analyse seulement les images distinctes et non pas les vidéos, il est possible que les détectations ne soient pas cohérentes entre deux images successives, et qu'un objet ne soit pas constamment détecté sur l'entière de son passage à la caméra (présence d'un débris, de détectations différentes entre les images ou détectations d'artefacts sur une seule image). La dernière partie du pipeline vise à remédier à ce problème, en sélectionnant uniquement les détectations cohérentes dans l'espace et le temps.

Figure 2 – Schématisation du pipeline développé pour analyser automatiquement les vidéos enregistrées par les caméras acoustiques.



1. Séquence d'étapes ou de processus automatisés mis en place pour traiter, transformer et analyser des données de manière séquentielle.
2. Contour rectangulaire tracé autour d'un objet ou d'une région d'intérêt dans une image.

### Suivi du passage des poissons détectés

À chaque détection du modèle, l'algorithme de post-traitement va comparer les coordonnées de la boîte englobante pour l'image donnée avec les coordonnées de la boîte englobante pour la détection sur l'image suivante. Le post-traitement va uniquement conserver les détections qui se succèdent logiquement, réduisant ainsi le nombre de détections de faux positifs, ou détections d'objets n'étant pas des poissons. L'opérateur peut paramétrer le nombre minimum de détections successives à conserver pour considérer la présence d'un poisson, et le nombre d'images vides autorisées entre les détections. Ces deux seuils sont donc de nouveaux paramètres qu'il convient d'étudier et d'optimiser car ils peuvent être propres à chaque site ou à chaque espèce cible. Ils sont par défaut fixés à cinq détections successives minimum en autorisant jusqu'à trois images vides entre deux détections.

L'identification de l'espèce du poisson par post-traitement correspondra à la classe majoritairement prédite par le modèle tout au long de sa détection. Les deux autres classes les plus probables sont conservées en deuxième et troisième probabilité de classification. De plus, le sens de déplacement des poissons détectés par le modèle est calculé à partir des coordonnées de la première et de la dernière détection d'un individu.

### Export des données

Au terme de l'analyse, une vidéo est recréée à partir des images brutes non prétraitées, en affichant les boîtes englobantes associées à chaque détection, ainsi que le label de la classe la plus probable tel qu'estimée par le réseau de neurones (et sa probabilité associée). Pour chacune des vidéos, est créé un tableur au format .csv décrivant les caractéristiques de chaque passage de poisson détecté.

### Résultats : les performances du pipeline

Les résultats du déploiement du réseau de neurones sur le jeu de validation mettent en évidence qu'il est possible de bien détecter les passages de poissons sur les images enregistrées par les caméras acoustiques, bien que les performances soient très dépendantes de la taille des individus (tableau 4).

Plus la taille du poisson est importante, plus le taux de détection par le modèle est élevé (cf. « Toutes les données », tableau 4). Les performances du modèle, satisfaisantes pour les individus de plus de 60 cm, sont nettement inférieures pour détecter les plus petits individus (tableau 4). Il est important de souligner que dans ce jeu de données, les poissons « génériques », appartenant à des espèces qui ne sont pas ciblées pour les questions de recherche ou dont l'espèce ne peut pas être formellement identifiée, sont majoritaires (cf. « Autres espèces », tableau 3). Les résultats montrent également que les poissons mesurant entre 40 et 80 cm sont moins bien détectés sur les vidéos enregistrées par la caméra ARIS que sur les images enregistrées par la caméra DIDSON, ce qui peut s'expliquer par la différence de taille de la fenêtre d'enregistrement (jusqu'à 16 m pour les vidéos ARIS vs. 10 m pour les vidéos DIDSON) : pour une même taille, l'écho d'un poisson représentera moins de pixels et sera donc plus difficile à détecter pour l'algorithme sur les données ARIS.

Les performances d'identification des espèces par le réseau sont évaluées sur les passages de saumons atlantiques (n = 65) et des anguilles européennes (n = 112). Sur les quarante heures de vidéos, les résultats présentés dans le tableau 5 montrent que 77 % des passages de saumons sont bien détectés, mais seulement 55 % des individus sont correctement identifiés comme des saumons sur les trois classes les plus probables (37 % sur la classe la plus probable). Les performances du réseau sont sensiblement identiques entre les vidéos DIDSON et ARIS pour cette espèce. Si l'efficacité de détection des anguilles est du même ordre que celle des saumons sur les données DIDSON (75 %), elle est nettement moindre sur les données ARIS (60 %) (tableau 5). La proportion bien plus importante d'anguilles d'une taille inférieure à 60 cm sur le jeu de validation ARIS (80 % des individus contre 45 % sur les données DIDSON, tableau 5) peut expliquer les moindres performances du réseau sur les données ARIS. Le réseau performe cependant de façon similaire pour l'identification des anguilles entre les données DIDSON et ARIS : respectivement 78 % et 80 % des anguilles détectées sont bien classées comme anguille sur les trois classes les plus probables.

Pour vérifier que le modèle ne détecte pas trop d'objets qui ne sont pas des poissons des espèces ciblées, et qui seraient donc visionnés à tort par un opérateur chargé de valider les prédictions du réseau, il faut considérer le nombre de faux positifs, ou individus mal classés. Pour les anguilles comme pour les saumons, les taux de faux positifs sont extrêmement élevés (tableau 5). Bien que

**Tableau 4 – Évolution du pourcentage de poissons bien détectés, en fonction de leur classe de taille et du modèle de caméra.**

Réseau	Taille du poisson (cm)				Total
	20-40	40-60	60-80	> 80	
Toutes les données	35,7 %	45,3 %	71,5 %	87,5 %	42,7 %
Données DIDSON	29,3 %	55,4 %	76,5 %	86,7 %	45,7 %
Données ARIS	43,7 %	33,8 %	62,8 %	88,9 %	38,9 %

**Tableau 5 – Identification des espèces d'intérêt dans le jeu de données de validation par le réseau de neurones en fonction du type de caméra.**

		Effectif total	Effectif détecté	Top3	Top2	Top1	Class. Top3	Faux positifs	Précision Top3
DIDSON	Saumon	53	41 (77 %)	29 (55 %)	27 (51 %)	19 (36 %)	824	795	3,5 %
	Anguille	32	24 (75 %)	17 (53 %)	14 (44 %)	11 (34 %)	680	663	2,5 %
ARIS	Saumon	12	9 (75 %)	7 (58 %)	6 (50 %)	5 (42 %)	301	294	2,3 %
	Anguille	80	48 (60 %)	39 (49 %)	39 (49 %)	29 (36 %)	603	564	6,5 %

Top1 : l'espèce de l'individu est évaluée comme la plus probable par le modèle ; Top2 : l'espèce de l'individu est évaluée comme l'une des deux classes les plus probables par le modèle ; Top3 : l'espèce de l'individu est évaluée comme l'une des trois classes les plus probables par le modèle ; Faux positifs : nombre d'individus classés de façon erronée comme saumons ou anguilles dans l'une des trois classes les plus probables par le modèle ; Précision Top3 : pourcentage d'individus correctement classés parmi les individus classés comme saumons ou anguilles dans l'une des trois classes les plus probables par le modèle.

les résultats montrent que les anguilles sont moins bien détectées sur les données ARIS, leur taux de vrais positifs est le plus important ici, avec 6,5 % des anguilles prédites en étant réellement (39 individus sur 603 anguilles prédites, tableau 5).

### Conclusions et perspectives

Le réseau de neurones entraîné dans le cadre de cette étude est capable de reconnaître des échos de poissons sur les images acoustiques, même isolées les unes des autres. L'ajout d'un algorithme assurant la cohérence temporelle entre les images favorise l'élimination de détections parasites et raffine les résultats du réseau pour mieux identifier les objets en mouvement. L'identification des espèces par le réseau de neurones est un challenge d'un niveau encore supérieur dont les résultats sont encourageants mais perfectibles. En effet, la validation par un opérateur des nombreux passages de poissons n'appartenant pas aux espèces ciblées reste toujours moins chronophage qu'une visualisation intégrale des chroniques vidéos. Cependant, le fort taux de faux positifs fait de cet outil une méthode, pour l'heure, seulement semi-automatique (Fernandez Garcia *et al.*, 2023), une validation par un opérateur restant une étape obligatoire pour s'assurer de la fiabilité des données quantitatives produites par le réseau. Pour améliorer significativement les performances du réseau, en termes de

taux de détection comme d'identification des espèces, il sera nécessaire d'inclure dès son entraînement les successions d'images issues du passage d'un même poisson pour que le modèle utilise le mode de nage de l'individu, son déplacement ou la déformation de son corps comme des critères de différenciation du bruit de fond ou d'une autre espèce, ce qu'une image fixe isolée ne permet pas de faire.

L'automatisation de la détection des passages de poissons amène également à la possibilité de collecter automatiquement des informations morphologiques importantes telles que la taille des individus par le biais d'algorithmes de segmentation sémantique<sup>3</sup> notamment. Ces informations morphométriques pourraient ensuite améliorer les performances générales d'identification spécifique, en ajoutant de l'expertise biologique (par le biais de classes de taille, de phénologie) à l'analyse informatique. Enfin, des algorithmes d'adaptation de domaine sont susceptibles de transposer la méthode développée dans le cadre de cette étude à d'autres sites de suivi, d'autres espèces et d'autres modèles de caméra (BlueView, Oculus notamment), dont les caractéristiques sont décrites dans l'encadré 1. ■

3. La segmentation sémantique est une tâche de vision par ordinateur qui s'appuie sur un algorithme d'apprentissage profond afin de classer chaque pixel en lui associant une étiquette.

## RÉFÉRENCES

- Ahmad, M., & Sundararajan, D. (1987). A fast algorithm for two dimensional median filtering. *IEEE Transactions on Circuits and Systems*, 34(11), 1364-1374. <https://doi.org/10.1109/TCS.1987.1086059>
- Christin, S., Hervet, É., Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10, 1632-1644. <https://doi.org/10.1111/2041-210X.13256>
- Dougherty, E. R., & Lotufo, R. A. (2003). *Hands-on morphological image processing (vol. 59)*. SPIE press, 290 p.
- Fernandez Garcia, G., Corpetti, T., Nevoux, M., Beaulaton, L., & Martignac, F. (2023). AcousticIA, a deep neural network for multi-species fish detection using multiple models of acoustic cameras. *Aquatic Ecology*, 57, 881-893. <https://doi.org/10.1007/s10452-023-10004-2>
- Jocher, G. (2020). Yolo network from ultralytics LLC.
- Lamba, A., Cassey, P., Segaran, R. R., & Koh, L. P. (2019). Deep learning for environmental conservation. *Current Biology*, 29(19), R977-R982. <https://doi.org/10.1016/j.cub.2019.08.016>
- LeCun, Y., Bengio, Y., & Hinton, G. E. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- Martignac, F., Daroux, A., Bagliniere, J. L., Ombredane, D., & Guillard, J. (2015). The use of acoustic cameras in shallow waters: new hydroacoustic tools for monitoring migratory fish population. A review of DIDSON technology. *Fish and Fisheries*, 16(3), 486-510. <https://doi.org/10.1111/faf.12071>
- O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy, Broadway Books.
- Wei, Y., Duan, Y., & An, D. (2022). Monitoring fish using imaging sonar: Capacity, challenges and future perspective. *Fish and Fisheries*, 23(6), 1347-1370. <https://doi.org/10.1111/faf.12693>
- Zivkovic, Z., & Van Der Heijden, F. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognition Letters*, 27(7), 773-780. <https://doi.org/10.1016/j.patrec.2005.11.005>