



HAL
open science

Automatisation De L'analyse De Données Des Caméras Acoustiques Par Réseaux De Neurones Convolutifs (CNN) Rapport d'étape

Guglielmo Fernandez Garcia, François Martignac, Thomas Corpetti, Marie
Nevoux, Laurent Beaulaton

► To cite this version:

Guglielmo Fernandez Garcia, François Martignac, Thomas Corpetti, Marie Nevoux, Laurent Beaulaton. Automatisation De L'analyse De Données Des Caméras Acoustiques Par Réseaux De Neurones Convolutifs (CNN) Rapport d'étape. [Rapport de recherche] OFB; INRAE; Institut Agro - Agrocampus Ouest; UPPA. 2020, 13 p. hal-03233535

HAL Id: hal-03233535

<https://hal.science/hal-03233535v1>

Submitted on 24 May 2021

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.

Automatisation De L'analyse De Données Des Caméras Acoustiques Par Réseaux De Neurones Convolutifs (CNN)

Rapport d'étape

Guglielmo FERNANDEZ GARCIA^{1,2}
François MARTIGNAC^{1,2}
Thomas CORPETTI³
Marie NEVOUX^{1,2}
Laurent BEAULATON^{1,2,4}

¹ UMR 0985 Ecologie et Santé des Ecosystèmes, INRAE-
Agrocampus Ouest

² Pôle Gestion des Migrateurs Amphihalins dans leur
Environnement, OFB, INRAE, AGROCAMPUS OUEST, UNIV PAU &
PAYS ADOUR/E2S UPPA

³ UMR 6554 Littoral, Environnement, Géomatique, Télédétection,
CNRS-Université Rennes 2

⁴ OFB, Direction de la recherche, de l'expertise et des données

Février 2020

• AUTEURS

Guglielmo FERNANDEZ GARCIA, chargé de recherche, contrat post-doctoral (UMR 0985 Ecologie et Santé des Ecosystèmes, INRAE-Agrocampus Ouest ; Pôle Gestion des Migrateurs Amphihalins dans leur Environnement, OFB, INRAE, AGROCAMPUS OUEST, UNIV PAU & PAYS ADOUR/E2S UPPA), guglielmo.fernandez-garcia@inrae.fr

François MARTIGNAC, ingénieur de recherche (UMR 0985 Ecologie et Santé des Ecosystèmes, INRAE-Agrocampus Ouest ; Pôle Gestion des Migrateurs Amphihalins dans leur Environnement, OFB, INRAE, AGROCAMPUS OUEST, UNIV PAU & PAYS ADOUR/E2S UPPA), francois.martignac@inrae.fr

Thomas CORPETTI, chargé de recherche (UMR 6554 Littoral, Environnement, Géomatique, Télédétection, CNRS-Université Rennes 2), thomas.corpetti@univ-rennes2.fr

Marie NEVOUX, chargée de recherche (UMR 0985 Ecologie et Santé des Ecosystèmes, INRAE-Agrocampus Ouest ; Pôle Gestion des Migrateurs Amphihalins dans leur Environnement, OFB, INRAE, AGROCAMPUS OUEST, UNIV PAU & PAYS ADOUR/E2S UPPA), marie.nevoux@inrae.fr

Laurent BEAULATON, chef de pôle (OFB, Pôle Gestion des Migrateurs Amphihalins dans leur Environnement, AFB, INRA, AGROCAMPUS OUEST, UNIV PAU & PAYS ADOUR/E2S UPPA), laurent.beaulaton@ofb.gouv.fr

Droits d'usage : restreint

Niveau géographique : national

Couverture géographique : France

Niveau de lecture : professionnels

- **AUTOMATISATION DE L'ANALYSE DE DONNÉES DES CAMÉRAS ACOUSTIQUES PAR RÉSEAUX DE NEURONES CONVOLUTIFS (CNN), FERNANDEZ GARCIA G., MARTIGNAC F., CORPETTI T., NEVOUX M., BEAULATON L.**

- **RÉSUMÉ**

L'UMR Ecologie et Santé des Ecosystèmes (INRAE-Agrocampus Ouest, Rennes) travaille depuis 2013 sur la description des migrations de poissons diadromes à l'aide de caméras acoustiques, en collaboration étroite avec l'Office Français de la Biodiversité dans le cadre du pôle OFB-INRAE-Agrocampus Ouest-UPPA pour la gestion des migrateurs amphihalins dans leur environnement. Ces sonars multi-faisceaux à très haute fréquence permettent d'obtenir des images de poissons sur lesquelles des caractéristiques morphologiques et comportementales peuvent être extraites. La mise en œuvre de ces outils sur les bords des cours d'eau permet une observation du comportement naturel des poissons sans structure porteuse lourde (barrage, seuil), même en eau turbide. Les caractéristiques des caméras acoustiques permettent l'acquisition de nouvelles données par les gestionnaires qui utilisent couramment des techniques d'imagerie pour suivre les populations de poissons migrateurs. À l'heure actuelle, sur une cinquantaine de sites en France, des caméras optiques installées sur des passes à poissons permettent de compter le nombre de poissons migrant chaque année. Le développement de nouveaux outils d'analyse de données issues des caméras acoustiques vise à faciliter l'utilisation de ces méthodes de suivi innovantes par les gestionnaires.

Pour l'heure, l'analyse des données enregistrées par les caméras acoustiques (DIDSON et ARIS, du constructeur Soundmetrics Corp.) passe par une visualisation intégrale ou partielle des images par des opérateurs, procédure fastidieuse et très coûteuse en temps. Les campagnes d'acquisition se déroulant sur plusieurs mois, une automatisation du traitement de ces images semble une avancée indispensable pour continuer à développer l'utilisation de ces outils en France, en particulier pour les transférer aux gestionnaires des milieux aquatiques. Les objectifs sont donc de mettre en place i) une détection automatique de poissons sur les enregistrements, et ii) une identification des espèces en fonction de critères morphologiques (taille, ondulation...) et comportementaux (direction, vitesse de passage...). Pour développer cette automatisation, des approches innovantes de traitement d'images seront mises en œuvre, via les réseaux de neurones convolutifs (CNN) notamment, en collaborant avec l'UMR Littoral Environnement Télédétection Géomatique (CNRS – Université de Rennes 2), laboratoire spécialisé en géographie physique et traitement mathématique d'imagerie complexe. Le développement de ces outils s'appuiera sur les données acquises dans le cadre d'études préalables du pôle migrateurs. A terme, l'optimisation des temps d'analyse des vidéos acoustiques élargira le nombre de stations pouvant être suivies ou allongera les campagnes d'acquisition. Pourront ainsi être abordées de nouvelles questions de recherche : dispersion d'une espèce diadrome sur un bassin versant en multipliant les sites suivis par des caméras acoustiques, étude de l'influence de l'évolution des conditions environnementales sur les phénologies de migration de plusieurs espèces diadromes qui fréquentent un même cours d'eau. Des applications pour la gestion peuvent également en résulter, dont la mise en œuvre temporaire des caméras acoustiques en complément d'une autre méthode de suivi des migrateurs pour estimer les flux de migration (projet Toolbox, 2016-2018).

- **MOTS CLÉS (THÉMATIQUE ET GÉOGRAPHIQUE)**

Réseau de Neurones Profonds ; Intelligence artificielle ; Caméra acoustique

• SOMMAIRE

I. Contexte et Objectifs	5
II. Les Réseaux de Neurones Profonds (CNN)	5
III. Approches méthodologiques	6
IV. Avancement 2019	6
IV.1. Définition de la stratégie Deep Learning	6
IV.2. Sélection et labellisation des données	7
IV.3. Débruitage et estimation du mouvement	8
V. Perspectives et objectifs 2020-2021	9
V.1. Développement de l'approche Deep Learning	9
V.2. Élaboration d'un ensemble de protocoles de travail pour le traitement des données	10
V.3. Développement du réseau des acteurs et élaboration des protocoles de travail	10
V.4. Perspectives futures 2021	11
V.5. Calendrier prévisionnel 2020-2021	11

- **AUTOMATISATION DE L'ANALYSE DE DONNÉES DES CAMÉRAS ACOUSTIQUES PAR RÉSEAUX DE NEURONES CONVOLUTIFS (CNN)**

I. 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 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.

Cette technologie intéresse donc un public de plus en plus larges 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 par exemple. Cependant, malgré leurs nets avantages, les principales limites de ces méthodes restent le volume des données enregistrées et l'analyse de ces vidéos, très chronophage en l'absence d'outils automatique de lecture des vidéos acoustiques. L'objectif de cette étude est d'appliquer une méthode innovante d'automatisation de l'analyse d'images aux données enregistrées par les caméras acoustiques : le réseau de neurones profonds.

II. Les Réseaux de Neurones Profonds (CNN)

Les algorithmes d'apprentissage profond permettent d'"apprendre" une distribution statistique connue (par exemple générée par un modèle physique ou par des observations) et d'en générer ou d'en classer une autre selon la distribution initiale. Au cours de la dernière décennie, ces algorithmes ont prouvé leur utilité dans de nombreux domaines, notamment en écologie. 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. D'un point de vue mathématique, l'apprentissage approfondi (ou Deep-Learning, DL) est un problème d'optimisation : il s'agit en pratique d'optimiser un grand nombre de paramètres liés aux fonctions "neurones", qui, empilés en "couches profondes", permettent de modéliser des phénomènes complexes. Chaque "couche profonde" représente une transformation effectuée sur les données d'intérêt ou, avec un langage plus proche de la vision par ordinateur, une convolution entre un filtre et les données. Ces opérations, en modifiant les données, génèrent donc des "descripteurs" des caractéristiques des données.

Il existe un nombre croissant de types de réseaux neuronaux, selon la structure différente des "neurones" et des "couches profondes" : les paramètres des "neurones" individuels 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.

III. Approches méthodologiques

D'un point de vue méthodologique, l'intérêt de ce projet ne se limite pas seulement à l'analyse efficace des populations avec les caméras acoustiques, mais aussi à l'élaboration d'un modèle efficace, simple et incluant des aspects physiques. Par ce dernier terme, il faut entendre l'inclusion de la physique du mouvement (cf. IV.1), pour limiter les aspects "empiriques" des réseaux de neurones. L'utilisation d'informations supplémentaires dans les données d'apprentissage permet de conserver une architecture de réseau simple et plus interprétable. Pour obtenir un tel modèle, il est cependant nécessaire de prétraiter les données et, en même temps, d'évaluer de manière critique l'architecture à utiliser.

IV. Avancement 2019

IV.1. Définition de la stratégie Deep Learning

Le panorama des méthodes d'apprentissage approfondi (ou Deep-Learning) est très large. Pour cette raison, il était nécessaire de définir la meilleure stratégie à suivre dans ce projet. Dans le domaine de l'analyse vidéo, deux approches peuvent être mises en œuvre : l'analyse de chaque image distincte d'une vidéo ou l'analyse de l'entièreté de l'action, en intégrant la corrélation spatio-temporelle de la vidéo. Ces dernières méthodes, bien qu'elles puissent sembler plus intéressantes pour la problématique de ce projet, présentent des limites non négligeables. Par exemple, il s'agit de réseaux de neurones qui nécessitent des ressources de calcul très importantes et il n'existe pas de critères de référence standard. En outre, aucun réseau neuronal de ce type n'est encore considéré comme « standard » et le sujet de la « reconnaissance de l'action » reste un débat ouvert dans la communauté scientifique.

Pour ces raisons, il a été décidé d'utiliser une stratégie différente, basée sur l'analyse de chacune des images des vidéos d'intérêt. Le réseau de neurones développé dans le cadre de ce projet est créé à partir de réseaux de neurones standards, dont l'efficacité est prouvée dans la communauté scientifique, adaptés à la problématique. Parmi ces réseaux, les deux plus intéressants sont le Faster R-CNN (Faster Region-based Convolutional Neural Networks¹), dont l'efficacité sera évaluée en priorité, et le réseau YOLOv3 (You Only Look Once)². Ces réseaux de neurones sont couramment utilisés pour la reconnaissance d'objets dans les données issues de caméras optiques et sont déjà entraînés sur de grands ensembles de données de référence et avec un grand nombre de classes (type

1 **Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks**, Ren S., He K., Girshick R. and Sun J., Advances in Neural Information Processing Systems (NIPS), 2015

2 **YOLOv3**, Redmon J., and Farhadi A., arXiv, 2015

d'objets).

Les images des caméras acoustiques montrent un très faible rapport signal/bruit : pour améliorer l'efficacité du modèle, il est donc nécessaire de prétraiter les images afin de mieux identifier le signal d'intérêt (voir IV.3). Même avec ce prétraitement, l'utilisation de réseaux neuronaux pour la reconnaissance d'une seule image n'est pas suffisante. En fait, de la même manière qu'une lecture par un opérateur, l'élément qui permet d'identifier le passage d'un individu est la corrélation spatio-temporelle entre les différentes images (ou frames). Cette corrélation sera récupérée par le calcul du mouvement entre les différentes images (cf. IV.3). En résumé, la stratégie définie est composée de l'analyse simultanée, à travers un réseau de neurones :

- des images des caméras sonar,
- des mêmes images suivant un traitement de débruitage,
- du mouvement des pixels de l'image par rapport aux images suivantes.

IV.2. Sélection et labellisation des données

Afin de développer les méthodes de DL mentionnées ci-dessus, il est cependant nécessaire de disposer d'un grand nombre d'images dans lesquelles les experts ont déjà identifié la présence (ou non) d'une espèce 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 que le modèle soit ne soit pas dépendant des données d'apprentissage, il est important que les images choisies reflètent la variabilité des situations auxquelles on souhaite l'appliquer.

En créant ces jeux de données, nous nous sommes d'abord attachés à identifier le type de variabilité qui nous intéresse : la caméra utilisée (ARIS ou DIDSON), les différentes espèces (saumon, anguille, etc.), les conditions de prise de vue (conditions de débits contrastées), les différentes profondeurs, la largeur spatiale de la prise de vue, la dissimulation du passage par rapport au fond et la vitesse du passage. Une fois que les vidéos d'intérêt ont été sélectionnées, les différentes images successives composant les vidéos en ont été extraites. Chacune des images est ensuite labellisée, c'est-à-dire que l'on dessine un rectangle (appelé "bounding box") dans lequel l'espèce est incluse, puis l'on indique quelle espèce est présente. Cette caractérisation crée un fichier d'annotation (format .txt ou .xml) pour chacune des images labellisées. Ces opérations sont réalisées à la fois par l'utilisation de scripts que nous avons produits et à la fois par des programmes disponibles en ligne gratuitement (e.g. LabelImg).

Les données sont ensuite organisées selon les formats standards utilisés dans la vision par ordinateur : en particulier Pascal VOC³, mais la conversion automatique au format YOLO et COCO⁴ est en cours.

Deux familles de jeux de données d'apprentissage ont été développées : les jeux de données à classe unique et les jeux de données à classes multiples. Dans la première, toutes les espèces sont identifiées comme une famille d'objets appelés "fish", tandis que dans la seconde, chaque espèce est identifiée par un code spécifique à trois lettres (par exemple ANG pour Anguille européenne, cf. Tableau 1), pour un total de neuf classes. Le premier type d'ensemble de données sera utilisé pour développer des modèles qui

³ **The PASCAL Visual Object Classes Challenge: A Retrospective**, Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A. *International Journal of Computer Vision*, 111(1), 98-136, 2015

⁴ **Microsoft COCO: Common Objects in Context**, Lin T.-Y., Maire M., Belongie S. J., Bourdev L. D., Girshick R. B., Hays J., Perona P., Ramanan D. and Zitnick C. L., *CoRR*, [abs/1405.0312](https://arxiv.org/abs/1405.0312), 2014

analysent le passage non spécifique d'une espèce, tandis que le second permettra de distinguer également les différentes espèces. Il est important de souligner la taille de ces jeux de données : les jeux de données de test sont constituées d'un minimum de 350 images étiquetées, tandis que ceux utilisés pour obtenir des résultats atteignent jusqu'à 6000 images étiquetées.

À l'heure actuelle, les jeux de données suivants ont donc été traités :

- «DIDSON_générique_test»: exclusivement des données DIDSON, classe unique, environ 400 images classées
- «ARIS_générique_test»: exclusivement des données ARIS, classe unique, environ 350 images classées
- «DIDSON_générique_large»: exclusivement des données DIDSON, classe unique, 3198 images classées
- «ARIS_générique_large»: exclusivement des données ARIS, classe unique, 1619 images classées
- «ALL_générique_test»: union de jeux de données «DIDSON_générique_test» et «ARIS_générique_test»
- «DIDSON_multi_large»: les mêmes images que « DIDSON_générique_large », mais labellisées avec les classes spécifiques des espèces d'intérêt (cf. Tableau 1)
- «ARIS_multi_large»: les mêmes images que «ARIS_générique_large », mais labellisées avec les classes spécifiques des espèces d'intérêt (cf. Tableau 1)
- «ALL_multi_large»: union de jeux de données «DIDSON_multi_large» et «ARIS_multi_large»

Tableau 1 : Codes spécifiques des espèces d'intérêt

Espèce	Code
Alose	ALO
Anguille européenne	ANG
Autre espèce	FISH
Brochet	BRO
Lamproie marine	LPM
Saumon atlantique	SAT
Silure glane	SIL
Truite de mer	TRM

IV.3. Débruitage et estimation du mouvement

Comme la vidéo de la caméra acoustique a un faible rapport signal/bruit, les ensembles de

données doivent être prétraités avant d'être utilisés pour apprendre les modèles DL. Nous nous sommes d'abord concentrés sur la soustraction de l'arrière-plan dans les vidéos : après avoir testé plusieurs méthodes, nous avons choisi d'utiliser un algorithme appelé « Gaussian Mixture-based Background/Foreground Segmentation » (les paramètres ont été optimisés pour les caméras ARIS et DIDSON). Avec ce traitement, les éléments fixes de la vidéo (fond) sont éliminés mais un fort bruit "sel et poivre" persiste. C'est un type de bruit caractérisé par un certain nombre de pixels à des valeurs extrêmes (0 ou 255 pour une image 8 bits), répartis dans l'espace de façon aléatoire sur l'image. Dans le cas d'un passage d'un petit individu ou d'un passage partiellement caché par la topographie du fond, ce bruit peut être comparable au signal donné par l'espèce ou masquer son passage : bien qu'elles fournissent des informations utiles, les images sans fond nécessitent un traitement supplémentaire, afin de distinguer le signal d'intérêt du bruit restant.

À partir des vidéos sans fond, diverses images ont ensuite été extraites et celles-ci composent un nouveau type de jeu de données, totalement identique à ceux introduits auparavant, mais avec le suffixe "_bkg" : dans ces jeux de données, sont donc considérés à la fois les images originales et celles sans fond.

Afin d'éliminer le bruit "sel et poivre", la "morphologie mathématique" a été utilisée. Sans entrer dans les détails de cette théorie, une série de filtres ont été optimisés afin de réduire le bruit, en affectant le moins possible le signal d'intérêt. Ce nouvel ensemble d'images, nommé avec le suffixe "_den", compose avec les images sans fond et les originaux, un autre nouveau type de jeu de données.

Lors de l'analyse manuelle des données, le mouvement de l'individu facilite sa détection par l'opérateur. Pour inclure cette corrélation temporelle entre les images, un algorithme de calcul de mouvement de type Lucas Kanadé⁵ a été développé : les images séquentielles sont examinées par paires et le déplacement de chaque pixel entre les images est calculé. Cet algorithme fournit des "cartes" de vecteurs bidimensionnels, où chaque vecteur constitue la vitesse d'un pixel donné. Ces données, associées aux images sur lesquelles elles sont calculées, composent les dernières informations qui seront utilisées pour la formation des réseaux de neurones. Lors de l'élaboration des scripts nécessaires pour exécuter les différentes étapes du traitement des images, un soin particulier a été apporté pour obtenir un processus aussi automatisé et simple que possible, du point de vue de l'utilisateur, l'objectif étant de faciliter la création d'une boîte à outils pour préparer le modèle aux développements futurs.

V. Perspectives et objectifs 2020-2021

Les objectifs pour l'année 2020 peuvent être répartis autour de trois axes de travail : (i) le développement de l'approche Deep Learning, (ii) l'élaboration d'un ensemble de protocoles de travail pour le traitement des données et (iii) le développement d'un réseau entre les acteurs impliqués dans les caméras acoustiques et l'élaboration des protocoles de travail afin d'envisager la livraison de la méthode d'analyse à des tiers.

V.1. Développement de l'approche Deep Learning

⁵ Lucas-Kanade 20 Years On: A Unifying Framework, Baker S., Matthews I., *International Journal of Computer Vision*, 56, 221–255, 2004

Le premier objectif de l'année 2020 sera le développement des réseaux neuronaux, à commencer par le Faster R-CNN. Seront testées, dans l'ordre, les données sans traitement, les données sans fond, les données débruitées et enfin l'intégralité des données précédentes avec le calcul du mouvement. Ce faisant, il sera nécessaire de modifier la structure des réseaux neuronaux afin de prendre en compte les différents types d'informations en même temps.

Grâce à la variabilité des ensembles de données développés, il sera possible d'étudier, par exemple, s'il est nécessaire de développer des modèles spécifiques à un seul type de caméra ou s'il est possible d'en développer un modèle générique.

Tout d'abord, un modèle de classe unique sera développé, capable de reconnaître le passage d'un individu mais pas de son espèce. Une fois que cela sera mis en œuvre, il sera possible de l'affiner en vue d'une reconnaissance spécifique. Compte tenu de la nature des images, on s'attend à des intervalles de confiance plus importants pour la reconnaissance de la présence d'un individu que pour une reconnaissance spécifique. En fonction des résultats obtenus, la nécessité d'affiner un deuxième modèle, séquentiel au premier, sera évaluée afin d'améliorer la reconnaissance de l'espèce mais aussi d'extraire des informations biologiques d'importance, telles que la taille de l'individu. En ce sens, il sera possible d'ajouter un modèle qui analyse les images, selon les méthodes de segmentation.

V.2. Élaboration d'un ensemble de protocoles de travail pour le traitement des données

Comme déjà mentionné au point IV.3, une attention particulière a déjà été accordée au développement de scripts de calcul simples afin de faciliter la diffusion de la méthode aux utilisateurs, membres de la communauté scientifique ou gestionnaires des milieux aquatiques. A ce sujet, les objectifs pour 2020 comprennent : (i) la rédaction d'un protocole de travail (sous forme de documentation) qui permet à un utilisateur d'effectuer le prétraitement des données pour l'apprentissage de manière autonome et (ii) la réécriture et le conditionnement des scripts afin qu'ils soient plus facilement exportables (via docker ou pyinstaller). L'objectif général est de fournir des outils simples (documentation et une série d'exécutables) qui permettent à un utilisateur de caméras acoustiques, pas nécessairement expérimenté en programmation, de préparer facilement des ensembles de données. Celles-ci peuvent ensuite être utilisées pour réemployer la méthode en cours de développement également sur des données de différents types, de différentes résolutions, etc. Cette nouvelle phase d'apprentissage permettra d'améliorer le modèle et de le rendre plus général. L'intérêt, en perspective, est de mettre en place un modèle le moins dépendant possible des données traitées pour la phase de formation. Dans ce cadre, les problèmes critiques qu'un utilisateur pourrait trouver dans le traitement des données ont déjà été identifiés et la réécriture des scénarios est déjà en cours: une première version non définitive du protocole a déjà été traitée et une deuxième version est prévue pour la fin février.

V.3. Développement du réseau des acteurs et élaboration des protocoles de travail

Le développement du point précédent servira également de point de départ pour une diffusion à un réseau national d'utilisateurs de caméras acoustiques qui pourraient être intéressés par l'analyse vidéo automatique. Ce réseau comprendra non seulement des acteurs universitaires, mais aussi des associations territoriales et l'établissement public territorial de bassin (EPTB). L'objectif est tout d'abord d'augmenter la variabilité des

données de formation afin de généraliser le modèle : chaque site de suivi présente des caractéristiques différentes, des espèces étudiées différentes, éventuellement des outils différents et une approche différente des instruments numériques.

Dans un deuxième temps, lorsque le modèle sera terminé, la boîte à outils sera déployée auprès des différents acteurs. Il est donc nécessaire de comprendre dès maintenant leurs besoins et de fournir les outils nécessaires pour permettre le déploiement futur du modèle (par exemple, la facilité d'utilisation des applications, la documentation, les tutoriels). Pour livrer nos protocoles à un tel réseau, il est également nécessaire d'établir des critères pour l'échantillonnage de données provenant d'un utilisateur tiers (la quantité, les caractéristiques, la variabilité des données, etc.) : les éléments pour la sélection des données, présentés en IV.2, devront être formalisés dans un document d'orientation, permettant à chaque utilisateur de sélectionner les données à utiliser, tout en restant dans les caractéristiques nécessaires au développement du modèle. Ce point sera développé tout au long de l'année 2020 pour assurer un développement rapide du réseau de contacts et du plan d'échantillonnage.

V.4. Perspectives futures 2021

Une fois que les points ci-dessus auront été développés, le modèle pourra entrer dans une phase de déploiement. Cela implique le développement d'une boîte à outils qui pourra être utilisée à l'avenir de façon autonome à chaque nouvelle installation, en commençant par les instruments déjà mentionnés dans la section précédente. Toujours en évaluant les besoins de la communauté, nous envisageons de développer une approche de virtualisation ou de conteneurisation.

En outre, nous prévoyons également d'interfacer le modèle directement avec les formats propriétaires (mais ouverts) utilisés dans les caméras sonar (par exemple les formats .ddf pour la technologie DIDSON ou .aris pour ARIS). Cela simplifiera grandement le déploiement et la maintenance future du modèle.

V.5. Calendrier prévisionnel 2020-2021

Janvier - Mars 2020 :

- Tests avec les réseaux Faster-RCNN sur les jeux de données génériques, avec et sans pré-traitement
- Tests avec les réseaux Faster-RCNN sur les jeux de données multi-espèces, sans pré-traitement
- Début développement sur le "plan d'échantillonnage"

Avril - Juin 2020 :

- Tests avec les réseaux Faster-RCNN sur les jeux de données multi-espèces et avec pré-traitement
- Développement du "plan d'échantillonnage" et début des démarches pour la récolte des données des usagers
- Déploiement local du modèle pour les données génériques

Juillet - Septembre 2020 :

- Tester les données des autres caméras (Oculus/Blueview)
- Déploiement local du modèle pour les données multi-espèces
- Évaluation et développement par rapport aux besoins supplémentaires (calcul de la

taille, etc.)

Octobre – Decembre 2020 :

- Finalisation protocoles de travail et articles
- Évaluer la stratégie pour la création d'une toolbox

Janvier – Avril 2021 :

- Développement de la toolbox

Avec le soutien financier de



ofb.gouv.fr/



www.inrae.fr/



www.agrocampus-ouest.fr



www.univ-pau.fr