







GonIAde : application de l'IA à des images d'échographies de gonades pour estimer la fécondité chez la truite commune

Mémoire de stage de fin d'études

Structure d'accueil : INRAE

Lieu de stage : 65 rue de Saint-Brieuc, 35 000, Rennes

Étudiante : Lorane De Piero

Maîtres de stage : François Martignac, Marie Nevoux, Jérôme Guitton, Marie-Pierre Étienne

Référent pédagogique : Sébastien Da Veiga

Promotion 2026

Remerciements

Je remercie tout d'abord mes quatre maîtres de stage, François Martignac, Marie Nevoux, Jérôme Guitton et Marie-Pierre Étienne, pour leur accompagnement et leurs conseils tout au long de ces six mois de stage. Je remercie également l'ensemble des personnes de l'INRAE et de l'Institut Agro pour leur accueil.

Je remercie également mon tuteur de stage, Sébastien Da Veiga, ainsi que l'ensemble du corps enseignant de l'ENSAI pour m'avoir accompagnée et formée pendant ces trois années.

Enfin, je remercie ma famille et mon copain pour leur soutien durant ce stage et, plus généralement, au cours de mes années à l'ENSAI.

Résumé

Ce rapport présente la problématique et les travaux effectués lors de mon stage de fin d'études à l'INRAE. L'objectif du stage était de développer un outil permettant d'aider les experts à estimer la fécondité des truites communes du bassin de l'Oir, une rivière suivie par l'INRAE. Pour cela, 1 204 échographies de truites réalisées entre 2019 et 2025, dont un sous-ensemble de 237 échographies a été annoté, ont été mises à disposition afin d'estimer le volume de leurs organes reproducteurs et de leurs œufs. Ces volumes permettent ensuite d'estimer la fécondité. Mon travail a ainsi consisté à appliquer et comparer différents modèles de segmentation d'images afin de segmenter automatiquement les organes reproducteurs et les œufs présents sur les échographies. Les segmentations obtenues ont ensuite été intégrées dans une chaîne de traitement permettant d'estimer les volumes puis la fécondité. Les modèles finaux retenus sont U-Net pour la segmentation des organes reproducteurs et YOLO pour celle des œufs. Les segmentations obtenues sont globalement satisfaisantes, malgré certaines échographies difficiles et une tendance de YOLO à surestimer le volume des œufs. Ces prédictions permettent de fournir aux experts une première segmentation qu'ils peuvent ensuite vérifier et corriger. Enfin, la chaîne de traitement développée permet d'obtenir une première estimation de la fécondité des truites étudiées, qui ne peut cependant pas être confirmée en l'absence de vérité terrain.

Abstract

This report presents the work carried out during my final-year internship at INRAE. The aim of the internship was to develop a tool to help experts estimate the fecundity of brown trout from the Oir basin, a river monitored by INRAE. For this purpose, 1,204 trout ultrasound images collected between 2019 and 2025 were made available, including a subset of 237 annotated images, in order to estimate the volume of the reproductive organs and the eggs. These volumes can then be used to estimate fecundity. My work consisted of applying and comparing different image segmentation models to automatically segment the reproductive organs and the eggs visible in the ultrasound images. The resulting segmentations were then integrated into a processing pipeline used to estimate the volumes and then the fecundity. The final models selected were U-Net for the segmentation of the reproductive organs and YOLO for the eggs. The resulting segmentations were generally satisfactory, although some ultrasound images remained difficult and YOLO showed a tendency to overestimate egg volume. These predictions provide experts with an initial segmentation that they can then check and correct. Finally, the developed processing pipeline provides a first estimate of the fecundity of the studied trout, although this estimate cannot be confirmed because no ground truth is available.

Mots-clés : truites communes, fécondité, échographies, segmentation d'images, réseaux de neurones

Table des matières

1 Introduction **1**

2 Environnement du stage **2**

3 La problématique et les données **3**
3.1 Enjeux et difficultés du problème 3
3.2 Acquisition et qualité des données 3
3.3 Description des données 4
3.3.1 Structure des données 4
3.3.2 Des truites de petite taille 5
3.3.3 Des échographies dépendantes du type et de la position de la coupe 5
3.4 Annotation des données 6
3.4.1 Une annotation difficile et incertaine 6
3.4.2 De nouvelles classes pour guider la segmentation 6
3.4.3 Annotation d'une partie des données 7
3.5 Prétraitement des données 8
3.5.1 Les images et les annotations 8
3.5.2 Les métadonnées 8

4 Méthodologie, expérimentations et résultats **9**
4.1 Environnement technique et reproductibilité 9
4.1.1 Ressources de calcul 9
4.1.2 Bibliothèques et frameworks 9
4.1.3 Reproductibilité des expérimentations 9
4.1.4 Rendre les méthodes accessibles à tous 9
4.2 Méthode d'estimation de la fécondité 10
4.2.1 Hypothèses pour l'estimation des volumes 10
4.2.2 Détail du calcul des volumes et de la fécondité 11
4.3 Méthodes de segmentation 12
4.3.1 État de l'art des modèles de segmentation d'images 12
4.3.2 Sélection et présentation des modèles 13
4.3.3 Post-traitement des prédictions 15
4.3.4 Métriques d'évaluation des performances des modèles 16
4.4 Entraînement, sélection et validation 17
4.4.1 Préparation des données d'apprentissage 17
4.4.2 Choix des hyperparamètres 17
4.4.3 Comparaison et sélection des modèles 18
4.5 Prédictions sur de nouvelles données 23
4.5.1 Performance des modèles retenus sur l'échantillon de test 23
4.5.2 Estimation de la fécondité 24
4.5.3 Vérification et utilisation des prédictions par les experts 26

5 Discussion **27**

6 Conclusion **30**

Annexes **31**

Table des figures

1 Illustration de différentes coupes transversales réalisées le long d'une truite 4
2 Schéma relationnel de la base de données associant les poissons aux échographies . . . 4
3 Échographies d'un même poisson pour différentes coupes et différentes positions . . . 5
4 Pipeline d'estimation de la fécondité à partir des échographies . . . 10
5 Schéma de l'approximation de la forme des gonades . . . 11
6 Distribution des scores Dice des gonades pour les différents modèles . . . 20
7 Visualisation des prédictions des modèles sur les échographies transversales 20
8 Visualisation des prédictions des modèles sur les échographies longitudinales 22
9 Dice U-Net pour les 24 échographies transversales de test (gonades) . . . 23
10 Dice YOLO pour les 11 échographies longitudinales de test (œufs) . . . 23
11 Visualisation des prédictions U-Net et YOLO sur des échographies transversales et longitudinales . . . 24
12 Comparaison du volume des gonades estimé à partir des prédictions et des annotations . . . 24
13 Comparaison du volume médian des œufs estimé à partir des prédictions et des annotations . . . 25
14 Relation entre la fécondité prédite et la fécondité calculée à partir des annotations . . . 25
15 Processus d'acquisition des échographies des truites sur le terrain . . . 31
16 Exemple d'une échographie avec plusieurs artefacts . . . 31
17 Schéma anatomique de la cavité d'une truite femelle : observation directe vs échographie . . . 32
18 Schéma simplifié de l'architecture de YOLO26-seg. . . . 32
19 Architecture U-Net classique, d'après Ronneberger et al. [3] . . . 33
20 Schéma simplifié de l'architecture CSA-Net. . . . 34
21 Principe général du modèle SegHVUNet . . . 35
22 Schéma simplifié du pré-entraînement d'OpenUS. . . . 36
23 Exemple de deux œufs fusionnés . . . 37
24 Séparation attendue des deux œufs . . . 37
25 Principe de la méthode de ligne de partage des eaux . . . 37
26 Exemple de prédictions avant et après application du post-traitement sur des échographies longitudinales . . . 38
27 Visualisation des prédictions YOLO26 avec l'ajout de la classe gonade . . . 38
28 Comparaison des scores Dice U-Net avec et sans CME pour chaque catégorie 39
29 Distribution des scores Dice des œufs . . . 41
30 Différence entre la surface médiane prédite et annotée des œufs avec YOLO26 41
31 Différence entre la surface médiane prédite et annotée des œufs avec U-Net 42
32 Différence entre la surface médiane prédite et annotée des œufs avec SegHVUNet . . . 42
33 Différence de surface entre les gonades prédites par U-Net et annotées . . . 42
34 Visualisation de l'interface sur les poissons de 2025 . . . 43
35 Visualisation de l'interface sur les poissons de 2025 . . . 43
36 Évolution des surfaces annotées des gonades selon la position des coupes 44
37 Exemple d'échographies à la position Op+2 . . . 44
38 Distribution des volumes des œufs estimés à partir des annotations . . . 45
39 Relation entre la fécondité estimée et la longueur du poisson mesurée . . . 45

Liste des tableaux

1 Échantillon de données sélectionné pour l'entrainement et la validation des modèles 7
2 Résultats obtenus par validation croisée sur les données transversales . . 19
3 Résultats de la validation croisée sur les échographies longitudinales . . . 21
4 Hyperparamètres des modèles 40
5 Métriques de rappel et précision des modèles 40

1 Introduction

Des chercheurs de l'INRAE [1] s'intéressent au suivi de la population de truites communes (*Salmo trutta*) présente dans le ruisseau de la Roche, affluent de l'Oir, en Normandie. La rivière et ses affluents font l'objet d'un suivi scientifique par l'INRAE et l'OFB [2] depuis près de quarante ans. Ces suivis permettent de comprendre pourquoi les populations augmentent, diminuent ou se maintiennent, et comment elles réagissent aux changements du milieu. Ils peuvent répondre à des enjeux de gestion, de conservation et de restauration des populations ou de leurs habitats.
Les données recueillies ont ainsi permis de développer des modèles de dynamique de population afin d'étudier l'évolution des populations de poissons migrateurs du bassin versant, en particulier celle de la truite commune. La fécondité, définie comme le nombre d'œufs par femelle, est l'un des paramètres nécessaires à ces modèles. Celle-ci étant difficile à obtenir directement sur des individus vivants, des méthodes d'échographie ont été mises en place afin d'obtenir une estimation non invasive de la fécondité. Cette estimation repose sur l'identification des gonades [3] et des œufs visibles sur les échographies.
Ainsi, depuis 2015, des échographies de truites sont réalisées lors de captures régulières par pêche électrique. Cependant, les échographies sont longues et coûteuses à analyser, et étaient jusqu'alors peu utilisées. L'échographie constitue en effet une donnée difficile à analyser en raison de son faible contraste, de la présence de bruit et d'artefacts, ainsi que de frontières anatomiques parfois peu distinctes.

Parallèlement, on observe depuis plusieurs années un développement important des méthodes d'intelligence artificielle appliquées à l'imagerie médicale. Ces approches sont notamment utilisées pour automatiser certaines tâches d'analyse d'images et aider les professionnels de santé dans leur interprétation. Plus particulièrement, l'utilisation de réseaux de neurones pour la segmentation d'échographies a connu des progrès importants. De nombreuses architectures et méthodes de segmentation ont ainsi été développées afin d'améliorer l'identification et la délimitation automatique des structures présentes dans ce type d'images.

Ce stage a ainsi pour objectif de répondre à la problématique suivante : comment appliquer des modèles d'intelligence artificielle à des images d'échographies de gonades pour estimer la fécondité chez la truite commune ? Cela se traduit par l'application, l'adaptation et la comparaison de modèles issus de la littérature afin de segmenter automatiquement les échographies de truites pour, à terme, estimer leur fécondité. Les modèles doivent notamment être capables de segmenter les gonades et les œufs présents sur les échographies. A partir de ces segmentations, un pipeline est mis en place pour estimer les volumes des gonades et des œufs. La combinaison de ces volumes permet alors d'obtenir une estimation du nombre d'œufs dans la gonade, et donc de la fécondité des femelles.

Ce rapport présente dans un premier temps le contexte et les enjeux du stage, les données utilisées ainsi que leur préparation et leur traitement. Dans un second temps, les méthodes de segmentation et d'estimation de la fécondité sont décrites. Enfin, l'évaluation des modèles, la qualité des estimations obtenues et les résultats finaux sont présentés et discutés dans la dernière partie.

1. Institut national de recherche pour l'agriculture, l'alimentation et l'environnement
2. L'Office Français de la Biodiversité
3. Organe reproducteur, correspondant à l'ovaire chez la femme.

2 Environnement du stage

J'ai effectué mon stage au sein de l'Unité Mixte de Recherche (UMR) DECOD "Dynamique et durabilité des écosystèmes : de la source à l'océan". Elle regroupe des équipes de recherche de l'INRAE, de l'Ifremer [4] et de l'Institut Agro, réparties sur les sites de Brest, Rennes, Lorient et Nantes. L'UMR DECOD a pour ambition de contribuer à l'avancement des connaissances afin d'anticiper les changements liés aux pressions exercées par les activités humaines sur les écosystèmes aquatiques. Ses activités sont centrées sur l'analyse des processus écologiques et évolutifs à l'œuvre dans les milieux aquatiques, l'évaluation de l'état de la biodiversité, et la construction de futurs scénarios possibles pour ces écosystèmes.

Mon stage a été réalisé en partie au sein de l'INRAE et en partie au sein du pôle Halieutique, mer et littoral de l'Institut Agro. Mes maîtres de stage sont rattachés à ces différentes structures. François Martignac et Marie Nevoux travaillent respectivement en tant qu'ingénieur de recherche et chargée de recherche à l'INRAE Rennes. Jérôme Guitton travaille en tant qu'ingénieur de recherche au sein du pôle halieutique de l'Institut Agro et Marie-Pierre Étienne en tant qu'enseignante-chercheuse à l'ENSAI (mais auparavant à l'Institut Agro). La variété des domaines de compétence de mes encadrants permet de combiner des approches complémentaires en écologie, en halieutique, en statistiques et en informatique afin de répondre aux problématiques du projet.

Mon stage s'inscrit dans un contexte de développement important de l'utilisation de capteurs dans la recherche et, parallèlement, de développement rapide des méthodes d'intelligence artificielle. De nombreux projets autour de l'IA sont ainsi mis en place pour accompagner les chercheurs dans le traitement et l'analyse de ces nouvelles données disponibles.
Dans le cas de ce projet, les échographies de truites, réalisées chaque année depuis 2015, étaient jusqu'alors très peu exploitées du fait des incertitudes liées à leur interprétation ainsi que du temps nécessaire pour extraire manuellement les informations pertinentes. Elles constituent cependant une source potentielle de données très intéressante à traiter et à analyser. L'automatisation de leur analyse à l'aide de méthodes d'intelligence artificielle vise ainsi à obtenir une première estimation de la fécondité des truites étudiées.

4. Institut Français de Recherche pour l'Exploitation de la mer

3 La problématique et les données

3.1 Enjeux et difficultés du problème

L'objectif du stage est de fournir un outil d'aide à l'évaluation de la fécondité des truites communes. La fécondité correspond ici au nombre d'œufs par femelle. Elle est estimée en faisant le ratio du volume des gonades (qui contiennent les œufs) sur le volume moyen d'un œuf. La fécondité est une variable très difficile à mesurer de manière non invasive. Une méthode pour l'estimer consiste à réaliser des échographies des truites afin de visualiser les organes reproducteurs (gonades) et les œufs qu'ils contiennent.

Cependant, les échographies prises dans le cadre de cette étude correspondent à des coupes 2D des poissons. Elles ne permettent donc pas de visualiser l'intégralité des gonades ni l'ensemble des œufs. Afin d'estimer la fécondité, des méthodes d'approximation des volumes des gonades et des œufs doivent ainsi être mises en place à partir des informations disponibles sur les échographies, notamment à partir des surfaces extraites sur les différentes coupes. Or, les échographies sont nombreuses et difficiles à lire en raison de leur faible résolution, de leur faible contraste et de la présence éventuelle de bruit et d'artefacts. Elles demandent ainsi beaucoup de temps ainsi qu'une forte expertise pour être analysées manuellement.

Le stage a donc deux objectifs : le premier est de fournir un outil capable d'aider les experts à segmenter les échographies, et le second est d'obtenir une première estimation de la fécondité des truites. Pour répondre à ces deux objectifs, la mission principale a consisté à appliquer des modèles d'IA aux images d'échographies afin de segmenter automatiquement les gonades et les œufs. Les experts peuvent alors utiliser les prédictions des modèles comme une aide à l'identification de ces structures sur de nouvelles échographies. Ces prédictions peuvent également être directement intégrées dans un pipeline afin de récupérer les surfaces des gonades et des œufs, d'estimer leurs volumes, puis d'en déduire une estimation de la fécondité. Pour cela, l'un des principaux enjeux était de développer un modèle généralisable, capable de segmenter le plus fidèlement possible de nouvelles échographies. Un travail préalable de préparation, de traitement et d'annotation des données a ainsi été nécessaire.

L'une des spécificités de ce projet est que la vérité terrain est inconnue. En effet, il n'est pas possible de vérifier directement les estimations obtenues sans recourir à des méthodes conduisant à la mort des individus analysés ou de leur ponte ("stripping" [5]), afin de compter le nombre réel d'œufs. Les performances des modèles de segmentation ont ainsi été évaluées à partir des annotations définies par les experts et utilisées comme références.

3.2 Acquisition et qualité des données

Les échographies sont réalisées sur le terrain depuis 2015, lors de pêches électriques organisées par l'INRAE sur le cours d'eau de la Roche, en Normandie, dans le bassin versant de l'Oir. Les pêches sont effectuées en octobre, juste avant la période de ponte des truites, qui commence en novembre.

La méthode d'acquisition des données d'échographies est une méthode non invasive, où la truite est anesthésiée avant chaque manipulation. Des mesures de poids, de taille et des prélèvements d'écailles ou génétiques sont également effectués. La truite, plongée dans l'eau, est ensuite passée sous l'échographe afin de déterminer son sexe et la longueur de

5. Technique qui consiste à faire pression sur l'abdomen du poisson pour faire sortir les œufs.

ses gonades. Comme illustré sur la figure 1, des coupes transversales sont ensuite réalisées tous les 2 cm environ, de l'opercule du poisson (début de la gonade identifiable) jusqu'à la fin de la gonade. La position des coupes est repérée à l'aide d'un mètre ruban attaché à la sonde et est essentielle pour permettre l'estimation ultérieure du volume des gonades (voir figure 15 en annexe). Une coupe longitudinale est également réalisée au début et/ou à la fin de la gonade, là où les œufs sont les moins denses.

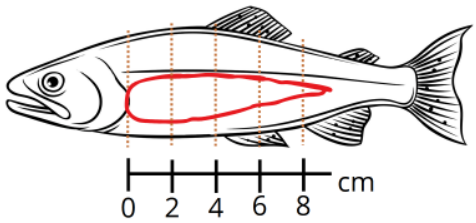


FIGURE 1 – Illustration de différentes coupes transversales réalisées le long d'une truite

L'obtention d'échographies de haute qualité est très difficile. La qualité des images dépend de nombreux facteurs et peut varier fortement d'une échographie à l'autre et d'un poisson à l'autre : position de la sonde, caractéristiques des tissus, présence d'artefacts. Par exemple, la présence de poches d'air ou de tissus durs (comme l'opercule) peut perturber la propagation des ultrasons et masquer une partie de l'image (voir figure 16 en annexe). Un même organe peut aussi présenter une échogénicité (capacité d'un organe à renvoyer les ultrasons) variable, comme c'est le cas des œufs en fonction de leur maturation. Par ailleurs, les poissons, bien qu'anesthésiés, peuvent bouger entre deux prises d'images, ce qui peut causer des erreurs de mesure et de précision dans le repérage spatial des coupes.

De plus, la qualité des données dépend également des moyens et du matériel utilisés. Par exemple, l'utilisation d'un appareil permettant de prendre des images 3D (actuellement échographies 2D) pourrait fournir des informations supplémentaires sur la continuité et le volume des gonades. Enfin, les truites étudiées sont des truites de rivière de petite taille. Leurs organes peuvent donc être petits et, par conséquent, plus difficiles à identifier que chez des truites plus grandes, issues de pisciculture par exemple.

3.3 Description des données

3.3.1 Structure des données

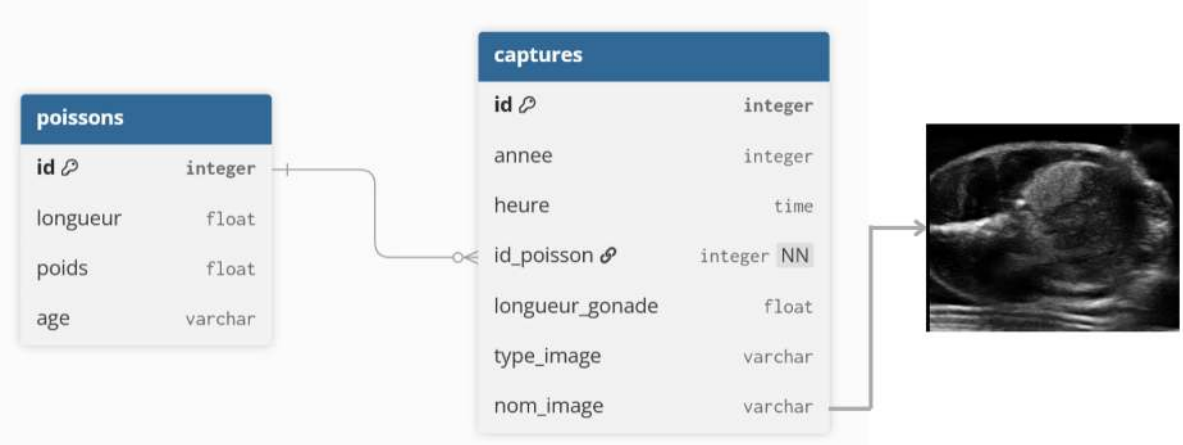


FIGURE 2 – Schéma relationnel de la base de données associant les poissons aux échographies

La structure des données est assez simple. Comme le montre la figure 2, les données textuelles sont organisées dans deux tables : une première table avec les caractéristiques des poissons (poids, âge et longueur) et une seconde table regroupant les métadonnées des captures (id, position, heure de capture, etc.). C'est cette seconde table qui permet de faire le lien entre la base d'images d'échographies (images JPEG de 640x480 pixels) et leurs métadonnées, à partir du nom de l'image. Les deux tables sont reliées entre elles par un identifiant unique attribué à chaque poisson.

3.3.2 Des truites de petite taille

Les poissons étudiés sont des truites communes de petite taille. Les truites analysées mesurent en moyenne 21 cm de longueur, avec des valeurs allant de 7 à 36 cm. Leurs gonades mesurent entre 2 et 14 cm. À titre de comparaison, les truites vivant dans de plus grandes rivières peuvent atteindre environ 60 cm.

3.3.3 Des échographies dépendantes du type et de la position de la coupe

La base de données contient des échographies faites lors des pêches de 2019 à 2025, soit **1 204** échographies correspondant à **243** poissons. Ces images sont divisées en deux types : 872 échographies transversales et 332 échographies longitudinales. Les échographies transversales permettent de visualiser une coupe complète de la cavité abdominale, contenant notamment les gonades. Les échographies longitudinales permettent un "zoom" sur les œufs. Les différentes coupes sont visibles sur la figure 3. Un poisson a entre 2 et 7 échographies transversales (3.6 en moyenne) et entre 1 et 3 longitudinales (1.4 en moyenne). Le nombre d'échographies transversales dépend notamment de la longueur de la gonade.

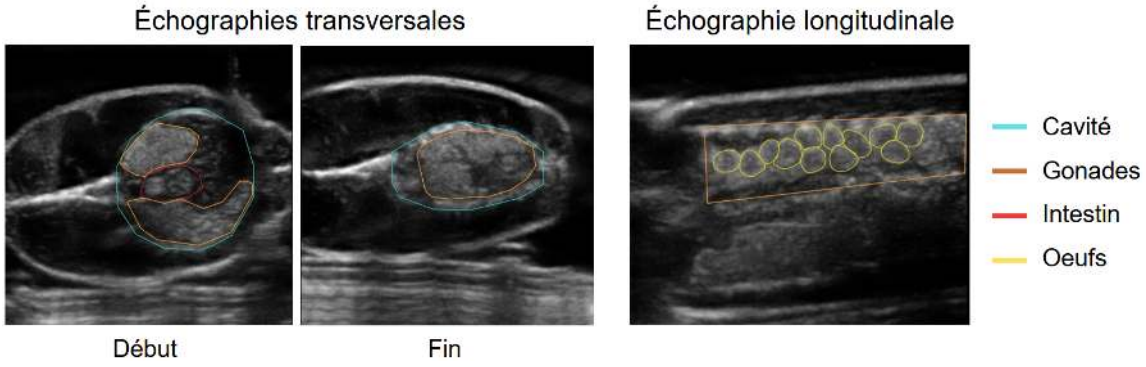


FIGURE 3 – Échographies d'un même poisson pour différentes coupes et différentes positions

Note : La première image montre la cavité générale sur une coupe transversale située vers la tête du poisson, avec les deux gonades distinctes entourant l'estomac. La deuxième montre la cavité générale vers la fin des gonades, où les deux gonades - indistinguables - occupent la majeure partie de la cavité. La troisième, correspondant à une coupe longitudinale, permet de visualiser plus précisément la gonade et les œufs qu'elle contient.

Une des particularités des échographies transversales est qu'elles présentent des différences importantes selon la position de la coupe. Comme illustré dans la figure 3, au début du poisson, les gonades sont contenues par d'autres organes, notamment le tube digestif et le foie (voir le schéma anatomique en annexe, figure 17). A l'inverse, vers la fin des gonades, les autres organes ne sont pas présents et les gonades occupent presque l'intégralité du volume de la cavité. La position de la coupe le long du poisson pourrait ainsi constituer une information supplémentaire pertinente pour la segmentation, puisqu'elle

permet notamment d'anticiper la présence de certains organes et leurs zones de recouvrement les plus probables. Outre la position, il existe également une certaine continuité anatomique entre les coupes successives d'un même poisson. Cette continuité (position, taille, échogénicité) pourrait ainsi être exploitée pour apporter un contexte supplémentaire à la segmentation des images.
Cependant, malgré ces régularités, la position, la taille et la forme des organes peuvent varier considérablement, même pour un même individu ou pour des positions équivalentes. Cette variabilité inter et intra-poisson est l'une des plus grandes difficultés liées à ce type d'échographies.

On peut également remarquer que, comme les échographies sont prises sur le côté gauche du poisson, la gonade droite est généralement mieux visible. Par ailleurs, dès que l'on avance dans le poisson, il n'est généralement pas possible de les différencier.

3.4 Annotation des données

Afin de permettre l'entraînement et l'évaluation des modèles, une phase d'annotation manuelle des échographies a été mise en place. Les images ont d'abord été annotées sur papier par une experte, puis reproduites numériquement par moi même à l'aide de l'outil Roboflow [1]. Les annotations ont ensuite été exportées aux formats COCO et YOLO.

3.4.1 Une annotation difficile et incertaine

La qualité des annotations est extrêmement importante pour les performances des modèles. Des annotations de bonne qualité peuvent significativement augmenter leur efficacité et la qualité de leur segmentation.

Cependant, les échographies à disposition présentent des structures aux délimitations ambiguës, avec des formes complexes et partiellement visibles, qui varient d'une image à l'autre. De nombreux artefacts acoustiques peuvent également être présents. Ainsi, même les experts ont des difficultés à annoter précisément les images, et de nombreuses incertitudes persistent. En effet, deux experts peuvent produire deux annotations différentes pour la même image. A cela s'ajoutent les biais de l'annotation humaine liés à la personne qui annote. Dans notre cas, une seule experte a annoté les images.
Enfin, il existe aussi des imprécisions liées au tracé manuel des contours, qui ne sont pas réalisés au pixel près.

Les annotations utilisées constituent donc des références pour l'entraînement et l'évaluation des modèles, mais ne peuvent pas être considérées comme une vérité terrain.

3.4.2 De nouvelles classes pour guider la segmentation

Les deux structures principales que l'on cherche à segmenter sont les gonades et les œufs, afin d'obtenir leur volume. Or, pour faciliter leur annotation, les experts s'appuient également sur les organes à proximité. Ces derniers ont donc été annotés numériquement afin de fournir au modèle des informations anatomiques supplémentaires susceptibles d'améliorer la segmentation. La cavité générale, délimitée par les arêtes, et le tube digestif ont ainsi été annotés sur les images transversales, et la gonade, sur les images longitudinales (présentés sur la figure 3). La cavité, pour les images transversales, et la gonade, pour les images longitudinales, fournissent notamment un contexte spatial permettant de mieux localiser les structures d'intérêt et de guider leur segmentation.

Pour les images transversales, une classe supplémentaire "Ignore gonade" a également été ajoutée. Elle correspond aux pixels pour lesquels la présence ou non d'une gonade est possible mais incertaine, notamment à cause d'artefacts ou d'autres organes qui masquent la zone concernée. Si l'on ne sait pas si une gonade est présente au niveau d'une certaine zone, plutôt que de la considérer comme du fond, on l'indique via cette classe. Ainsi, au lieu d'entraîner le modèle à prédire du "background" à cet endroit, il sera entraîné à ignorer ces zones lors de l'entraînement en les excluant de la fonction de perte. L'idée est de ne pas imposer au modèle des informations potentiellement erronées.

Concernant les données longitudinales, seuls les œufs les plus visibles et les plus entiers possibles ont été annotés. L'objectif n'est donc pas de détecter l'ensemble des œufs présents, mais d'en identifier quelques-uns, de la bonne forme et de la bonne taille, puisque la fécondité se calcule à partir du rapport entre le volume des gonades et le volume de référence des œufs d'un poisson.

3.4.3 Annotation d'une partie des données

La segmentation manuelle des images est une tâche longue et coûteuse, d'autant plus lorsqu'elle nécessite une expertise spécifique. Bien qu'il s'agisse du scénario idéal, il était ainsi impossible d'annoter l'intégralité des données à disposition. Seul un échantillon a été retenu afin d'obtenir un équilibre entre une base d'apprentissage suffisamment importante et le temps consacré à l'annotation. L'échantillon a été tiré aléatoirement parmi quatre des sept années disponibles. Pour chaque année sélectionnée, les poissons ont été tirés aléatoirement, de manière uniforme et sans remise, jusqu'à atteindre un total de 40 échographies transversales. Les échographies longitudinales des poissons sélectionnés ont ensuite été récupérées. Au total, **237** échographies provenant de **49** truites ont été retenues pour la phase d'annotation. Les truites sélectionnées ont une taille proche de celle de la population, avec une longueur moyenne de 21 cm. La répartition des données sélectionnées par année est présentée dans la table 1 ci-dessous.

TABLE 1 – Échantillon de données sélectionné pour l'entrainement et la validation des modèles

Années	2019	2020	2022	2024	Total
Poissons	13	12	12	12	49
Échos transversales	43	40	43	43	169
Échos longitudinales	18	15	20	15	68

Pour donner un ordre d'idée, une image nécessite en moyenne 3 min de labellisation, soit près de 12 h pour l'ensemble des images sélectionnées, contre 60 h si les 1 204 images avaient été annotées. A cela s'ajoute un temps de vérification pour d'éventuelles corrections des annotations numériques. Pour gagner du temps, les organes secondaires (cavité et intestin sur les transversales, gonades sur les longitudinales) ont été annotés de façon "bruitée" avec des polygones plus simples et des contours moins précis.

Finalement, parmi les 68 échographies longitudinales retenues, aucun œuf n'a pu être annoté sur 10 d'entre elles, en raison de leur petite taille ou de la qualité insuffisante des images. Au total, 381 œufs ont été annotés, avec une moyenne de 6.5 œufs par échographie. Concernant les 169 échographies transversales, 180 gonades ont pu être annotées et 5 images ont été supposées sans gonades. De plus, 20 échographies ont été segmentées avec la classe "Ignore gonade", c'est-à-dire avec une gonade potentiellement incomplète.

3.5 Prétraitement des données

3.5.1 Les images et les annotations

Plusieurs prétraitements ont été appliqués aux échographies. Les transformations géométriques appliquées aux images ont également été appliquées aux masques [6] afin de préserver leur correspondance. D'autres traitements spécifiques leur ont été appliqués, notamment pour adapter l'organisation des classes aux formats attendus par les modèles.

Recadrage et redimensionnement
Les bordures des échographies, qui contiennent principalement des informations techniques de l'échographe et qui n'apportent pas d'informations utiles à la segmentation, ont été supprimées. Seule l'échelle de l'image a été conservée pour permettre la conversion des pixels en centimètres. Les dimensions des images sont donc passées de 640x480 à 510x380 pixels. Un padding a ensuite été ajouté pour obtenir une taille d'entrée d'image classique sans déformer la géométrie des structures : 512x512 pixels.

Normalisation
Les valeurs des pixels ont été normalisées entre 0 et 1 pour faciliter et stabiliser l'apprentissage des modèles. Les images sont au format RGB pour correspondre au format attendu par les modèles.

Débruitage
Les échographies présentent couramment ce qu'on appelle du bruit de "speckle". Des outils de débruitage ont été testés, mais sans effet notable, probablement en raison de la nature structurée du bruit de speckle, qui le rend difficile à dissocier des informations anatomiques utiles. De plus, ces méthodes peuvent lisser les contours des structures et supprimer des informations anatomiques utiles à la segmentation.

3.5.2 Les métadonnées

Les informations relatives aux poissons et aux échographies ont été regroupées dans la même base de données, permettant notamment de conserver uniquement les informations sur les échographies de truites femelles. Les données manquantes ont été supprimées et les erreurs identifiées ont été corrigées. Par exemple, il est arrivé que la position de la dernière échographie dépasse la longueur totale de la gonade, à cause d'un décalage lors de l'acquisition des coupes. Dans ce cas, la position des coupes a été réajustée afin que la dernière échographie corresponde à l'extrémité de la gonade.

Afin de caractériser la position des échographies le long des gonades, une nouvelle variable a été créée à partir de leur position relative. Les positions sont renseignées selon leur distance (en cm) à l'opercule, noté "Op". Les premières échographies du poisson, à "Op0" ou "Op+1", ont été classées dans la catégorie "début". Parmi les échographies restantes, celles dont la position relative était inférieure à 60% ont été classées dans la catégorie "milieu", tandis que les autres ont été classées dans la catégorie "fin". Cette classification permet de distinguer trois régions des gonades en fonction des positions des coupes.
De plus, une variable pour l'échelle de l'échographie a été ajoutée. Elle a été récupérée via OCR grâce à l'échelle indiquée en bas à droite de chaque échographie (package Tesseract). Elle correspond à la hauteur en cm de l'écran de l'échographe et permet de convertir les mesures réalisées en pixels en cm.

6. Représentations des structures annotées sur une image, stockées ici sous forme de coordonnées.

4 Méthodologie, expérimentations et résultats

Cette partie présente d'abord l'environnement technique et la méthode utilisée pour estimer la fécondité à partir des échographies. Les modèles de segmentation, leur entraînement et leur protocole d'évaluation sont ensuite décrits et comparés. Enfin, les modèles retenus sont appliqués au jeu de test puis intégrés dans l'interface destinée aux experts.

4.1 Environnement technique et reproductibilité

4.1.1 Ressources de calcul

Implémenter des modèles de deep learning nécessite des ressources de calcul importantes, notamment en termes de puissance de calcul et de mémoire. Ces contraintes doivent donc être prises en compte lorsque l'on veut entraîner des modèles en des temps acceptables.
Les expérimentations ont été réalisées sur un serveur distant sous Ubuntu 26.04, équipé de deux cartes graphiques NVIDIA RTX A5000 et d'un processeur Intel Xeon Silver 4214R. Les ressources de calcul étant partagées avec d'autres utilisateurs, leur utilisation a nécessité une gestion adaptée, notamment en surveillant la consommation de mémoire et en se coordonnant avec les autres utilisateurs.

4.1.2 Bibliothèques et frameworks

Le développement des différentes méthodes a été réalisé en Python, principalement à l'aide des frameworks PyTorch et Ultralytics, utilisés pour l'implémentation, l'entraînement et l'évaluation des modèles de segmentation.
Un environnement virtuel Python (version 3.14) a été créé et les bibliothèques nécessaires et leurs versions ont été regroupées dans un fichier `requirements.txt`.

4.1.3 Reproductibilité des expérimentations

Le code source du projet est disponible sur GitHub. Les configurations, les hyperparamètres et les métriques des modèles ont été sauvegardés sur le dépôt. Les données et les poids des modèles entraînés ont été conservés sur une plateforme de stockage à distance.
La reproductibilité du code est importante lorsque l'on veut comparer plusieurs modèles. Afin d'obtenir des résultats comparables, les graines des générateurs pseudo-aléatoires ont été fixées afin de contrôler notamment l'initialisation des modèles, la division des données et les potentielles transformations aléatoires. Des paramètres déterministes ont également été activés pour les opérations GPU.

4.1.4 Rendre les méthodes accessibles à tous

J'ai mis en place une interface graphique interne à l'aide de Streamlit afin de faciliter l'utilisation des modèles et des résultats par les experts. Elle permet de visualiser et de corriger, directement sur l'interface, les prédictions des modèles de segmentation sur de nouvelles échographies. Cela permet un gain de temps pour l'expert qui peut facilement s'appuyer sur les prédictions pour réaliser ses annotations. L'interface permet également d'obtenir directement, dans un fichier CSV, les estimations des volumes et de la fécondité de nouveaux poissons. Elle constitue ainsi un outil de restitution et de suivi des travaux réalisés pendant le stage, permettant de faciliter leur utilisation par les experts sans avoir à lancer directement le code.

4.2 Méthode d'estimation de la fécondité

La fécondité d'un poisson est estimée en divisant le volume total des gonades par le volume de référence d'un œuf. L'ensemble du pipeline, des échographies prétraitées jusqu'à l'estimation finale de la fécondité, est présenté sur la figure 4.

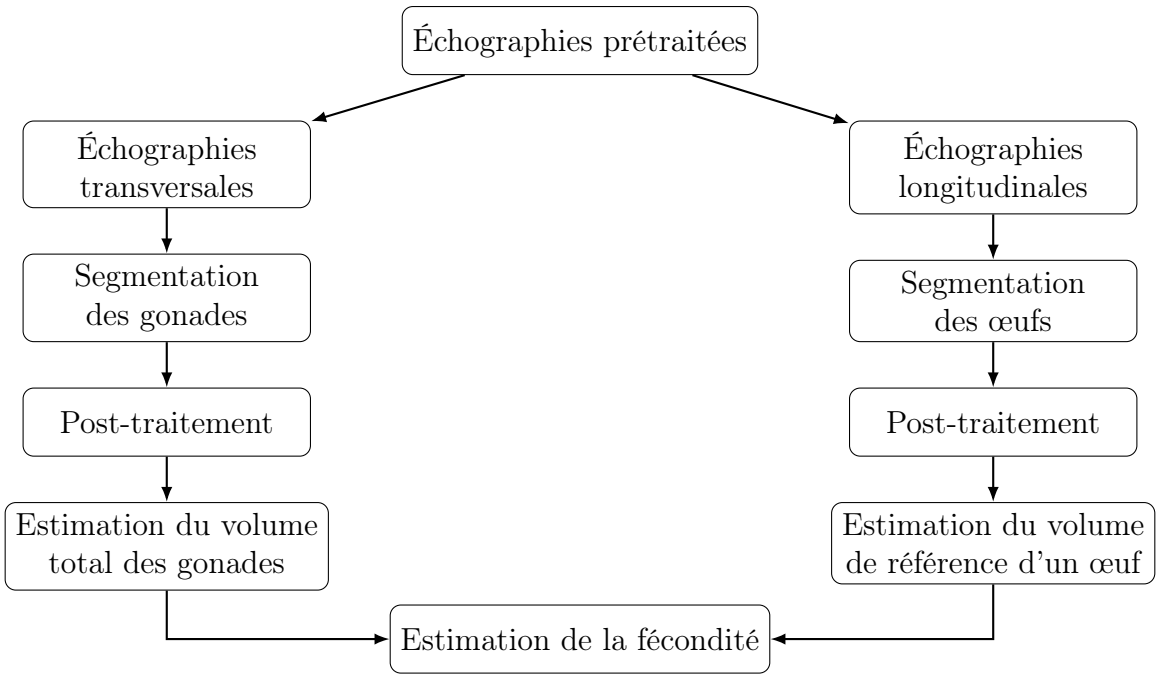


FIGURE 4 – Pipeline d'estimation de la fécondité à partir des échographies

La suite de cette section présente les hypothèses nécessaires à l'estimation des volumes, puis détaille les différentes étapes de calcul permettant d'obtenir une estimation de la fécondité à partir des segmentations produites par les modèles.

4.2.1 Hypothèses pour l'estimation des volumes

L'estimation des volumes repose sur plusieurs hypothèses :

- Les œufs d'un même poisson font tous la même taille biologiquement parlant ;
- Les échographies sont prises perpendiculairement à l'axe du corps du poisson ;
- La gonade et la cavité commencent au niveau de l'opercule, c'est-à-dire à "Op0" ;
- Les œufs sont supposés sphériques ;

Pour estimer le volume des gonades, on fait également l'hypothèse que leur forme globale est un tube conique, de sorte qu'une coupe transversale est approximée par un disque (figure 5). Ainsi, la forme des gonades correspond à une succession de cônes tronqués, prolongée par un cône jusqu'à la fin de la gonade. Un cône peut également être estimé au début de l'opercule, si la première échographie n'est pas présente.

A noter que, souvent, les gonades gauches et droites ne sont pas distinguables entre elles et le calcul est donc effectué comme s'il n'y avait qu'une seule grande gonade. De plus, en pratique, les gonades se développent autour d'autres organes déjà présents, ce qui leur donne des formes plus complexes que celles supposées.

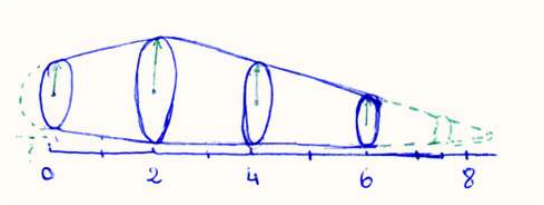


FIGURE 5 – Schéma de l'approximation de la forme des gonades

Note : Les cercles bleus correspondent aux surfaces des gonades sur les coupes disponibles.

Enfin, bien que les œufs d'un même poisson soient supposés faire tous la même taille, des variations de surface et de volume entre les œufs annotés d'un même poisson ont été observées. Ces variations peuvent être dues à des erreurs d'annotation mais également à des œufs à des distances différentes de la sonde, à des œufs coupés non perpendiculairement au centre de l'œuf ou à des œufs partiellement cachés. Afin d'obtenir un estimateur robuste du volume d'un œuf et limiter les valeurs extrêmes, le volume de référence d'un œuf est estimé par la médiane des volumes observés.

4.2.2 Détail du calcul des volumes et de la fécondité

Détail de l'estimation de la fécondité des truites pour un poisson donné :

1. Récupérer les coordonnées des masques des gonades et des œufs
2. Calcul de la surface en pixels des masques via la formule de shoelace :

$$A_{pixel} = \frac{1}{2}\sum_{i=1}^{n}(x_i y_{i+1} - x_{i+1} y_i),$$

   avec x_i et y_i les coordonnées horizontale et verticale du i-ème point définissant le contour du polygone.
3. Conversion de l'aire en cm^2 via l'échelle de l'image : $A_{cm^2} = (\frac{\text{echelle}}{\text{hauteur}})^2 * A_{pixel}$
4. Calcul du rayon du cercle correspondant à la surface calculée : $r = \sqrt{(A_{cm^2}/\pi)}$
5. **Œufs :**
   (a) Calcul du volume en cm^3 de l'œuf i, approximé par une sphère $V_i = \frac{4}{3}\pi r_i^3$
   (b) Estimer le volume des œufs du poisson par la médiane, qui est un estimateur robuste face aux œufs surestimés ou sous-estimés :

$$V_{\text{oeufs}} = \text{Med}(V_1, \ldots, V_n) \tag{1}$$

6. **Gonades :**
   (a) Calcul du volume en cm^3 des cônes tronqués :

$$V_i = h_i * \frac{\pi}{3} * (r_i^2 + r_{i+1}^2 + r_i r_{i+1}), \quad i = 1, \ldots, n-1,$$

   avec h_i la distance entre l'échographie i et i+1, r_i et r_{i+1} les rayons des bases.
   (b) Calcul du volume en cm^3 du cône final : $V_n = \frac{1}{3}\pi r_n^2 h_n$, avec r_n le rayon de la n-ième échographie et h_n la distance entre la n-ième échographie et la fin de la gonade.
   (c) Calcul du volume total de la k-ième gonade : $V_k = \sum_{i=1}^{n} V_i$ et somme des K gonades : $V_{gonades} = \sum_{k=1}^{K} V_k$

7. Calcul du **nombre d'œufs** (i.e. de la fécondité) : $F = \frac{V_{\text{gonades}}}{V_{\text{oeufs}}}$

Il est important de noter que la fécondité réelle n'est pas connue et qu'il n'est pas possible de vérifier la méthode de calcul proposée. En revanche, on sait qu'elle comporte des limites, notamment dans ses hypothèses de formes très simplifiées.

4.3 Méthodes de segmentation

4.3.1 État de l'art des modèles de segmentation d'images

Une recherche bibliographique des modèles de segmentation utilisés dans la littérature récente a été réalisée afin d'identifier les principales approches susceptibles de répondre à la problématique.

En segmentation d'images médicales, et plus particulièrement d'échographies, les architectures en U occupent une place importante dans la littérature [2]. Ces architectures sont tirées du célèbre U-Net de Ronneberger et al. en 2015 [3]. Parmi ces variantes les plus connues, on retrouve Attention U-Net, ResUNet et UNet++, qui introduisent notamment des mécanismes d'attention, des connexions résiduelles ou des connexions plus denses entre l'encodeur et le décodeur. D'autres, plus récentes, continuent à être régulièrement publiées, telles que ResAttentionUNet++ de Singh et al. en 2025 [4]. Des modèles portant sur des problématiques similaires aux nôtres ont également été développés. Yari et al. développent en 2024 des modèles U-Net modifiés combinant une tête de segmentation et de classification afin d'automatiser le suivi de maturation des ovaires du saumon atlantique [5] et d'évaluer la maturation des œufs [6].

D'autres approches sont également utilisées pour la segmentation d'images. Les modèles YOLO (You Only Look Once), initialement développés pour la détection d'objets, intègrent désormais des fonctionnalités de segmentation d'instances permettant de distinguer individuellement plusieurs objets d'une même classe. Parallèlement, les Vision Transformers (ViT) [7] exploitent des mécanismes d'attention pour modéliser les relations entre différentes régions d'une image. Les modèles de fondation, tels que SAM (Segment Anything Model), constituent une autre approche, permettant notamment de segmenter des objets à partir d'un prompt textuel ou visuel.

Cependant, les modèles pré-entraînés sur des données naturelles[7], comme ImageNet, ne sont pas forcément adaptés à la segmentation d'images spécifiques comme les échographies. Pour contrer cette limite, plusieurs modèles pré-entraînés spécifiquement sur des données médicales ou d'échographies ont été développés. La majorité de ces modèles sont développés sur la base du modèle SAM, tels que UniUltra de Li et al. en 2025 [8] et UltraSam de Meyer et al. en 2025 [9]. D'autres architectures existent, comme OpenUS de Zheng et al. en 2025 [10], basé sur une architecture Vision Mamba et un apprentissage auto-supervisé.

Un autre point intéressant à explorer en lien avec notre problématique est la dépendance séquentielle des échographies associées au même poisson. En effet, les modèles 2D classiques traitent chaque échographie indépendamment, sans exploiter les relations entre les coupes successives d'un même poisson. Les approches 2,5D, comme CSA-Net de Kumar et al. [11], permettent d'intégrer ces dépendances sans recourir à une architecture 3D.

7. Elles correspondent à des images du quotidien, comme des objets, des animaux ou des scènes.

Enfin, un dernier axe de recherche est d'examiner les liens selon la position de l'échographie dans le poisson (début, milieu ou fin). Une méthode pour tenir compte des différents types de coupe, sans réaliser plusieurs modèles spécifiques, est d'utiliser l'embedding des métadonnées comme canal RGB [12].

4.3.2 Sélection et présentation des modèles

Au final, 6 modèles ont été retenus et comparés durant ce stage, représentant différents types d'architectures.

YOLO
Le modèle choisi est YOLO26, publié par Ultralytics en 2026 [13], qui appartient à la famille des modèles YOLO, conçus initialement pour la détection d'objets en temps réel. L'un des grands avantages de YOLO est qu'il propose de la segmentation d'instance qui consiste à segmenter individuellement les objets d'une même classe. La segmentation d'instance est particulièrement pertinente pour notre objectif de segmentation individuelle des œufs, notamment lorsque plusieurs œufs sont proches ou en contact. C'est ce modèle qui sera appliqué à nos données comme référence pour la segmentation d'instance. Une explication générale de la structure du modèle est présentée dans l'annexe B.1.

U-Net
Le modèle U-Net [3] fait ici office de modèle de référence pour la segmentation sémantique. Très largement utilisé en imagerie médicale, il repose sur une structure encodeur-décodeur en forme de U. Son architecture est détaillée dans l'annexe B.2.
Dans notre étude, l'encodeur classique est remplacé par un ResNet-34 pré-entraîné sur ImageNet. Les connexions résiduelles facilitent l'entraînement du réseau, et le pré-entraînement permet de bénéficier de caractéristiques visuelles déjà apprises.

ResAttentionUNet++
Le modèle de Singh et al. [4] permet d'étudier une variante d'une architecture U-Net afin d'améliorer la segmentation des images d'échographies. Il propose plusieurs ajouts pour améliorer l'extraction des caractéristiques, faciliter leur fusion et mieux identifier les régions pertinentes de l'image.

Son architecture repose sur trois composantes principales : un ResNet-34 pré-entraîné sur ImageNet, un décodeur inspiré de UNet++ et un mécanisme d'attention par canal et spatial. Le rôle de chacune de ces composantes est détaillé dans l'annexe B.3. Il a été retenu pour l'étude afin de voir l'impact d'une architecture U-Net plus élaborée sur la segmentation des structures anatomiques.

CSA-Net
Cross-Slice Attention Network (CSA-Net) de Kumar et al. [11] est un modèle de segmentation 2.5D qui permet d'exploiter les dépendances anatomiques entre les coupes d'un même individu. Il constitue un compromis entre un modèle 2D et un modèle 3D très complexe. Dans notre étude, il permet de ne plus considérer les coupes successives d'un même poisson comme complètement indépendantes et de tirer parti de leur contexte spatial. Il est appliqué uniquement aux images transversales.

Le modèle prend en entrée trois coupes consécutives d'un poisson, dont les caractéristiques sont extraites via un ResNet partagé. Des mécanismes d'attention *Cross-Slice*

permettent ensuite d'apprendre les correspondances des régions entre les coupes voisines et celles de la coupe centrale. Un mécanisme d'attention *In-Slice* analyse ensuite les relations entre les différentes régions de la coupe centrale. Toutes ces informations sont finalement exploitées pour obtenir la segmentation de la coupe centrale. Le fonctionnement plus détaillé de l'architecture est présenté dans l'annexe B.4.

CSA-Net pourrait donc améliorer la segmentation des structures difficiles à identifier sur une image isolée en utilisant celles de ses voisins. Cependant, son architecture est bien plus complexe et coûteuse qu'un U-Net classique. De plus, le modèle original nécessite deux coupes voisines pour prédire la coupe centrale. Une adaptation a donc dû être réalisée pour que le modèle puisse segmenter les premières et dernières coupes des poissons. Enfin, la qualité de la prédiction dépend également de celle des coupes voisines. La présence de bruit, d'artefacts ou d'incohérences entre les images peut perturber les mécanismes d'attention et réduire les bénéfices de l'information inter-coupes.

SegHVUNet
L'objectif du modèle de Yari et al. [6] est d'évaluer la maturation des œufs. Ils ont pour cela développé un modèle issu d'une architecture U-Net capable de mieux distinguer les œufs individuellement. Dans notre étude, il est appliqué uniquement aux coupes longitudinales, afin de séparer individuellement les œufs des truites.

La particularité du modèle est qu'il prédit, en plus d'un masque de segmentation binaire, deux cartes horizontale et verticale (H/V). Ces cartes représentent la position relative de chaque pixel par rapport au centre de l'œuf auquel il appartient. Elles sont ensuite utilisées pour détecter les frontières entre les œufs adjacents grâce à un filtre qui permet de détecter les contours d'un objet (filtre de Sobel). Enfin, les frontières sont combinées au masque binaire pour obtenir des instances individuelles des œufs (cf annexe B.5). Pour leur application, une tête de classification supplémentaire permet également de prédire le stade de maturation des œufs. Elle n'a pas été conservée ici.

Le modèle permet de mieux détecter les objets compacts et proches les uns des autres, ce qui est souvent le cas des œufs. Il permet ainsi de limiter la fusion de masques, courante en segmentation sémantique, et donc de limiter les erreurs de surface dues à la fusion de plusieurs œufs. De plus, sa problématique étant directement applicable à la nôtre, le modèle a pu être implémenté tel quel. Il présente toutefois plusieurs limites : il donne de meilleurs résultats quand tous les contours des œufs sont annotés, et la séparation dépend de la qualité des cartes prédites.

OpenUS
L'objectif est de pouvoir étudier l'intérêt d'un modèle pré-entraîné sur des échographies plutôt que sur des images naturelles. Ce type de modèle pourrait permettre de mieux capturer les particularités des échographies, notamment lorsque, comme ici, les données annotées sont limitées.
OpenUS, proposé par Zheng et al. en 2025 [10], est un modèle qui utilise une architecture Vision Mamba capable de capturer à la fois des relations spatiales locales et à longue distance. Son pré-entraînement auto-supervisé combine apprentissage contrastif et reconstruction de régions masquées. Un masquage auto-adaptatif permet également de cibler les régions importantes et difficiles à reconstruire. Son fonctionnement plus détaillé est présenté dans l'annexe B.6.

Ce modèle a été retenu pour plusieurs raisons. Déjà, contrairement aux modèles pré-entraînés dérivés de SAM, comme UltraSam, il ne nécessite pas d'ajouter des indications supplémentaires à l'image, telles que des points ou des boites. De plus, un des grands avantages d'OpenUS est que les auteurs ont cherché à développer un modèle capable d'être transposé à différentes régions anatomiques et à différents appareils et protocoles. Pour cela, le modèle a été pré-entraîné sur un vaste corpus diversifié d'images échographiques (plus de 300 000 images échographiques, 42 jeux de données et 12 organes différents). Cette diversité des données de prétraitement est particulièrement intéressante pour nous, puisque les échographies de truites sont très spécifiques et ne ressemblent à aucune autre dans le milieu médical.
Un autre avantage, plus pratique, est que l'ensemble du code a été mis à disposition avec les poids pré-entraînés. Néanmoins, des adaptations au code initial ont dû être nécessaires pour l'appliquer aux données. Le modèle, implémenté initialement pour marcher avec des masques binaires, a dû être adapté pour qu'il accepte le multilabel. Cela a aussi entraîné, par conséquent, l'implémentation de métriques par classe. D'autres changements ont également été effectués pour harmoniser le protocole avec les autres modèles (cross-validation et data augmentation). De plus, OpenUS possède une architecture assez complexe, qui rend son adaptation et son entraînement plus difficiles. Enfin, malgré la diversité de son pré-entraînement, les échographies de truites restent très différentes de celles utilisées par les auteurs.

4.3.3 Post-traitement des prédictions

La quasi-totalité des modèles présentés est ce qu'on appelle "pixel-wise" (pour les modèles sémantiques), c'est-à-dire qu'une classe est attribuée de manière indépendante à chaque pixel de l'image. Les modèles ne regardent donc pas la prédiction comme une forme globale, mais comme un regroupement de pixels. Le problème avec ces méthodes est que les formes peuvent être très loin de celles d'origine, voire même impossibles biologiquement parlant. Les masques prédits peuvent par exemple avoir des contours irréguliers, des trous ou des petits îlots séparés du reste.
Pour aider à corriger ces problèmes, on peut appliquer un post-traitement sur les prédictions des modèles. Les opérations communément appliquées sont celles issues de la morphologie mathématique.

Opérations :

- **Seuillage par taille :** seules les composantes connexes dont la surface dépasse un seuil minimum sont conservées : 1000 pixels pour les gonades et 100 pour les œufs [8].
- **Limite du nombre de composantes :** une seule composante connexe est conservée pour la cavité et deux au maximum sont conservées pour les gonades.
- **Fermeture :** permet de combler les petits creux et irrégularités présents dans les contours.
- **Ouverture :** permet de supprimer les petites excroissances ou irrégularités tout en conservant la forme globale de l'objet (voir explication en annexe C.1).
- **Comblement des trous :** suppression des trous pour toutes les composantes.
- **Convexité :** une enveloppe convexe est appliquée sur la cavité qui conserve une forme assez stable entre les échographies, à l'inverse des autres structures.

8. L'équivalent en cm^2 change selon l'échelle de l'image, mais correspond à environ $0.1cm^2$ pour les gonades et $0.01cm^2$ pour les œufs, pour une hauteur fixée à 380 pixels.

Ces transformations ont en réalité un impact marginal sur les performances globales des modèles, mais elles permettent une bien meilleure cohérence biologique et visuelle des prédictions. Des exemples avant et après post-traitement sont présentés dans la figure 26, en annexe.

Une des problématiques des œufs, lorsqu'ils sont prédits par des modèles sémantiques classiques, était également leur séparation. Le post-traitement applique donc également une séparation des œufs qui se touchent selon la méthode "watershed". La méthode de séparation des œufs est détaillée dans l'annexe C.2.

4.3.4 Métriques d'évaluation des performances des modèles

Plusieurs métriques ont été utilisées dans cette étude. Les métriques classiquement utilisées en segmentation d'images telles que le coefficient de Dice et l'IoU (Intersection over Union) ont été implémentées. Ces deux métriques permettent de mesurer le recouvrement entre le masque prédit et le masque annoté. L'IoU étant une mesure plus stricte que le coefficient de Dice car il pénalise davantage les zones mal segmentées. Un coefficient ou un IoU de 1 signifie que les masques se superposent parfaitement. Le Dice correspond par définition au F1-score. Le rappel et la précision sont également utilisés pour obtenir une indication supplémentaire sur la segmentation des masques. Les métriques sont calculées de manière binaire, classe par classe.

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \qquad \text{IoU} = \frac{TP}{TP + FP + FN}$$

$$\text{Rappel} = \frac{TP}{TP + FN} \qquad \text{Precision} = \frac{TP}{TP + FP},$$

TP : Nombre de pixels correctement prédits comme appartenant à la classe ;
FP : Nombre de pixels prédits dans la classe alors qu'ils n'y appartiennent pas ;
FN : Nombre de pixels appartenant réellement à la classe mais non détectés.

Cependant, une des limites des métriques classiques est qu'elles ne parviennent pas toujours à capturer le contexte et la cohérence médicale ou biologique. Elles peuvent donner de très bons scores même si les masques obtenus ne parviennent pas à préserver la forme globale de l'objet. Elles sont donc souvent, dans le milieu médical, combinées à l'avis des experts pour la qualité des résultats. Dans notre cas, ces métriques ne sont pas adaptées aux œufs, qui, ne pouvant être distingués un à un individuellement (car cela aurait nécessité un appariement préalable entre instances prédites et réelles), sont évalués comme un même bloc. Or, dans notre cas, le fait que le modèle oublie ou ajoute des œufs n'est pas forcément un problème, tant qu'il en détecte un nombre suffisant. Dans cette situation, d'autres métriques plus spécifiques sont mieux adaptées à notre problématique, qui consiste à obtenir une bonne estimation de la taille moyenne des œufs afin d'estimer leur nombre.
La différence de surface entre le masque réel et le masque prédit a ainsi été calculée. Pour les échographies transversales, les différents masques de la même classe ont été regroupés et la différence totale de surface a été exprimée en cm^2. Pour les images longitudinales, ce qui nous intéresse, c'est de comparer le volume moyen individuel des œufs, et non le volume de tous les œufs regroupés. Pour cela, on prend la surface médiane, qui sera considérée comme la surface de référence des œufs de l'échographie. On obtient donc une différence de surface en mm^2 entre la surface médiane des œufs annotés et prédits.
Les zones annotées avec la classe "Ignore gonade" ont été exclues du calcul des métriques.

4.4 Entraînement, sélection et validation

4.4.1 Préparation des données d'apprentissage

Plusieurs choix ont été faits sur les données afin d'optimiser l'entraînement et la généralisation des modèles.

Choix des classes et des données d'entrée
L'annotation d'organes supplémentaires permet de donner plus de contexte anatomique aux modèles, mais augmente également le nombre de classes à apprendre. Plusieurs configurations ont donc été réalisées afin d'évaluer leur impact pour la segmentation des structures d'intérêt.
Deux classes ont finalement été retenues pour les échographies transversales (la cavité et la gonade) et une seule classe pour les échographies longitudinales (les œufs). En effet, les modèles ont eu beaucoup de mal à segmenter le tube digestif, qui n'apportait finalement pas d'améliorations à la segmentation des gonades. L'ajout de la gonade pour les échographies longitudinales n'a également pas permis de contenir et améliorer la segmentation des œufs (figure 27 en annexe).
L'utilisation des métadonnées pour améliorer la performance des modèles a également été étudiée, via une approche de Conditional Metadata Embedding (CME). La méthode consiste à encoder la position de l'échographie (début, milieu ou fin) sous un nouveau canal RGB supplémentaire ajouté aux images d'entrée. Ce choix a été motivé par l'observation de différences anatomiques entre les coupes situées au début, au milieu et à la fin des gonades. Cependant, les résultats obtenus (figure 28 en annexe) ne montrent pas une amélioration suffisante pour conserver les informations supplémentaires.

Augmentation des données
La data augmentation consiste à "augmenter" le jeu de données d'apprentissage sans annoter de nouvelles images en transformant les images existantes. L'objectif est de permettre au modèle de voir artificiellement plus d'images différentes et donc d'améliorer sa généralisation. Elle est particulièrement utile dans des situations où le jeu de données est relativement petit. Elle n'a pas été appliquée une seule fois, pendant le prétraitement, mais de manière dynamique à chaque nouvelle epoch [9], ce qui évite le stockage inutile des données et permet au modèle de voir une nouvelle version transformée de l'image à chaque epoch.
Les transformations appliquées sont la rotation ($\pm 20°$), la translation ($\pm 10\%$), la modification de l'échelle ($\pm 20\%$) et la modification de la luminosité (± 0.4 d'amplitude). Les transformations sont tirées aléatoirement dans l'intervalle donné, reproductibles via une graine fixée pour la comparaison entre modèles. Il faut faire attention à ce que les transformations effectuées soient toujours "réalistes" pour ne pas obtenir l'effet inverse de celui recherché. Les transformations géométriques ont également été appliquées aux annotations pour ne pas induire d'erreurs.

4.4.2 Choix des hyperparamètres

Les hyperparamètres choisis peuvent directement affecter les performances des modèles. Par exemple, un taux d'apprentissage trop faible peut ralentir la convergence du modèle, là où un taux d'apprentissage trop élevé peut rendre l'entrainement instable.
Pour les modèles reposant sur une architecture de type U-Net, les principaux hyperparamètres ont été définis à partir des données et des performances observées sur l'échantillon

9. Correspond à un passage complet de l'ensemble des données d'entraînement dans le modèle.

de validation. Pour YOLO et OpenUS, les paramètres d'entraînement proposés par défaut ont été conservés dans leur ensemble, avec quelques adaptations nécessaires aux expérimentations.

La fonction de perte retenue pour les variantes U-Net combine la Binary Cross Entropy (BCE) et la Dice loss (équations 3 et 4 en annexe) :

$$\mathcal{L} = \lambda \mathcal{L}_{\text{BCE}} + (1 - \lambda)\mathcal{L}_{\text{Dice}}, \qquad \lambda = 0.5 \tag{2}$$

L'association de ces deux fonctions de perte est couramment utilisée en segmentation médicale : la BCE pénalise les erreurs de prédiction à l'échelle des pixels, tandis que la Dice loss favorise le recouvrement entre les masques prédits et les annotations. Pour SegHVUNet, comme dans Yari et al. [6], une perte MSE (équation 5 en annexe) est ajoutée pour superviser la prédiction des cartes horizontale et verticale (H/V). OpenUS utilise une Binary Cross Entropy simple. YOLO utilise quant à lui la fonction de perte composite proposée par Ultralytics pour la segmentation d'instances.
Pour les modèles U-Net, les éventuelles zones annotées avec la classe "Ignore gonade" ont été exclues du calcul de la perte afin de ne pas pénaliser le modèle sur des régions dont l'annotation est incertaine.

L'optimiseur retenu pour les architectures dérivées de U-Net est AdamW. Il repose sur un optimiseur Adam avec une décroissance des poids (*weight decay*). OpenUS utilise quant à lui un optimiseur Adam. Les taux d'apprentissage initiaux ont été fixés à $10^{-}3$. Pour U-Net, il est réduit d'un facteur 10 après un certain nombre d'epochs consécutives sans amélioration de la perte de validation (*patience scheduler*). Cela permet de réduire progressivement l'amplitude des mises à jour lorsque l'apprentissage stagne. Pour YOLO, l'optimiseur est sélectionné automatiquement par Ultralytics.
Le nombre maximal d'epochs a été défini à 100 ou 150 selon les modèles. Pour les modèles U-Net et YOLO, un critère d'arrêt (*patience arrêt*) a été mis en place pour interrompre l'entraînement lorsque la perte de validation ne s'améliore plus pendant un certain nombre d'epochs. Cela permet notamment de limiter le temps de calcul et de réduire le risque de sur-apprentissage. Enfin, la taille des batchs, limitée par la VRAM du GPU, a été adaptée à chaque modèle. Les hyperparamètres retenus pour les différents modèles sont présentés dans la table 4 en annexe.

Les seuils de prédiction ont également été sélectionnés sur les données de validation. L'objectif était d'obtenir, pour chaque modèle, un compromis entre la précision et le rappel afin de limiter à la fois les faux positifs et faux négatifs. Les valeurs de précision et de rappel obtenues pour les différents modèles sont présentées en annexe table 5.

4.4.3 Comparaison et sélection des modèles

Les données ont été préalablement divisées en un ensemble de test (environ 15%) et un ensemble destiné à l'entraînement et à la validation des modèles. Ce dernier a ensuite été réparti selon une validation croisée 5-fold (via GroupKFold). La contrainte ici est que toutes les échographies d'un même poisson doivent appartenir au même ensemble, afin de pouvoir calculer les volumes et la fécondité par poisson. Les mêmes partitions ont été utilisées pour chaque modèle afin de garantir des conditions de comparaison identiques.

Les différents modèles ont été entraînés et évalués sur les 5 folds de la validation croisée. Leur comparaison se fait sur leur capacité à segmenter les gonades et les

œufs sur les ensembles de validation. Du fait que les deux tâches de segmentation - gonades et œufs - soient très différentes l'une de l'autre, les mêmes modèles ne sont pas forcément appliqués et comparés pour les échographies transversales et longitudinales.

Les gonades

La table 2 présente les performances moyennes obtenues sur les cinq folds de validation croisée pour la segmentation des gonades. Les scores de Dice et d'IoU sont très proches entre U-Net, ResAttentionUNet++ et OpenUS, avec des moyennes similaires au regard de la variabilité observée entre les folds. CSA-Net présente en revanche des performances moyennes inférieures à celles des autres modèles. Cela laisse donc supposer que, dans ce contexte, l'information issue des coupes successives n'améliore pas la segmentation. Elle pourrait même la perturber, par exemple à cause de la qualité des images très changeante d'une coupe à l'autre.

Concernant la différence moyenne de surface des gonades, U-Net présente les valeurs les plus faibles, avec environ 0.18 cm^2, suivi de ResAttentionUNet++ avec environ 0.20 cm^2, puis OpenUS avec 0.25 cm^2. Le décalage observé pour OpenUS entre les résultats obtenus pour les performances et la différence de surface montre qu'un bon recouvrement des masques n'entraîne pas systématiquement une meilleure estimation des surfaces, qui est également très importante pour notre problématique.

Un autre point intéressant dans la table 2 est le très grand écart entre le temps d'exécution des modèles. U-Net et ResAttentionUNet++ nécessitent environ 20 min d'entraînement, contre près du triple, soit environ 60 minutes, pour CSA-Net et OpenUS. Cette différence peut notamment s'expliquer par des architectures plus complexes et par l'utilisation de modèles pré-entraînés plus lourds. CSA-Net et OpenUS présentent ainsi un coût de calcul nettement supérieur, sans offrir d'améliorations dans les résultats moyens obtenus.

Méthodes	Temps	Dice score	IoU Score	Diff surfaces (cm^2)
U-Net	22 min	0.793 (± 0.054)	**0.691** (± 0.055)	**0.182** (± 0.043)
ResAttentionUNet++	16 min	**0.795** (± 0.047)	0.689 (± 0.052)	0.199 (± 0.058)
CSA-Net	60 min	0.731 (± 0.056)	0.609 (± 0.056)	0.265 (± 0.030)
OpenUS	60 min	0.794 (± 0.036)	0.689 (± 0.037)	0.246 (± 0.083)

Les résultats correspondent à la moyenne des 5-folds. L'écart type entre les folds est donné entre parenthèses.

TABLE 2 – Résultats obtenus par validation croisée sur les données transversales

La figure 6 apporte une analyse complémentaire en présentant la dispersion des scores Dice entre les échographies. On peut voir que tous les modèles présentent plusieurs valeurs extrêmes, correspondant à des images mal prédites. L'analyse de ces valeurs extrêmes montre qu'il s'agit presque toujours des mêmes échographies, et qu'elles correspondent donc à des images particulièrement difficiles à segmenter, tous modèles confondus. Deux images, en plus des cinq sans gonades annotées, présentent notamment des performances particulièrement faibles pour tous les modèles. Il s'agit d'échographies avec des œufs sur-matures, dont l'apparence est très différente de celle des œufs habituellement observés (noirs et très grands). Il y a très peu d'œufs sur-matures dans les données, ce qui rend leur segmentation d'autant plus difficile. Un exemple d'une des échographies est présenté sur la deuxième ligne de la figure 7.

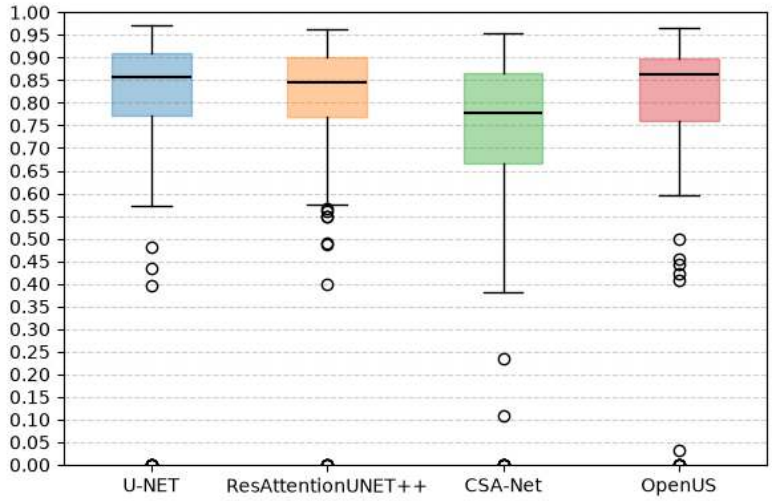


FIGURE 6 – Distribution des scores Dice des gonades pour les différents modèles
Données : échantillon de validation.

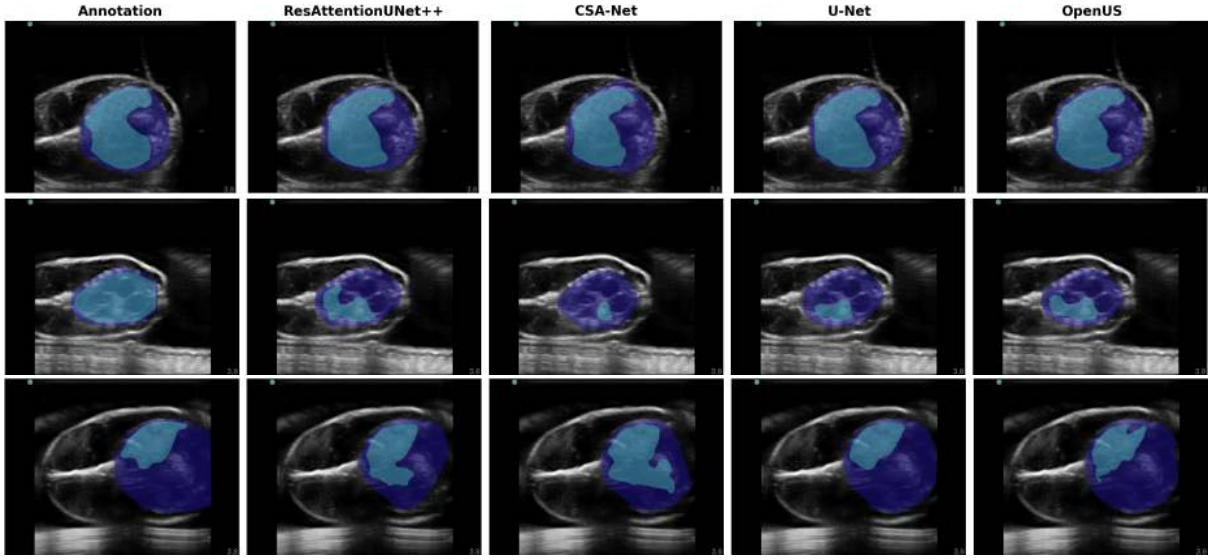


FIGURE 7 – Visualisation des prédictions des modèles sur les échographies transversales
Note : L'échographie de la première ligne présente une gonade bien segmentée avec un Dice supérieur à 0.9 pour tous les modèles. L'échographie de la deuxième ligne a des œufs sur-matures, avec une gonade mal segmentée pour tous les modèles. La troisième montre une gonade relativement bien segmentée pour U-Net et OpenUS mais moins bien pour CSA-Net et ResAttentionUNet++.

A l'inverse, certaines échographies sont très bien segmentées par l'ensemble des modèles, avec un Dice supérieur à 0.9 pour chacun des modèles. Il s'agit notamment d'images sur lesquelles les contours des gonades sont bien visibles et facilement identifiables, comme illustré sur la première ligne de la figure 7.
Enfin, la troisième ligne de la figure 7 présente un cas intermédiaire, pour lequel la qualité des prédictions varie d'un modèle à l'autre, sans qu'une architecture se distingue systématiquement sur l'ensemble des images. Sur l'image présentée, le modèle U-Net offre la meilleure segmentation avec un Dice de 0.89.

Ainsi, les deux modèles les plus adaptés, lorsque l'on considère les performances de segmentation, les différences de surfaces et le temps d'exécution des modèles, sont U-Net et ResAttentionUNet++. Le modèle final conservé est U-Net, qui minimise, sur les

données évaluées, la différence de surface.

Les œufs
La table 3 présente les résultats obtenus sur les données de validation. Le modèle SegHVUNet obtient un score Dice moyen supérieur à celui des autres modèles, bien que la différence reste faible au regard de la variabilité observée entre les folds. La figure 29 en annexe montre en effet la superposition des distributions des scores Dice pour les trois modèles. On observe également une superposition des différences de surfaces moyennes. Tous les modèles présentent des temps d'exécution du même ordre de grandeur. On peut également remarquer, à titre indicatif, que sur les 326 œufs annotés, YOLO26 est le modèle dont le nombre total d'œufs prédits est le plus proche, avec 320 œufs.
Cependant, toutes ces métriques agrégées ne permettent pas, à elles seules, d'évaluer les performances des modèles pour notre tâche de segmentation, et des analyses plus précises sont nécessaires.

Méthodes	Temps	Dice score	Diff surface (mm^2)	Nombre d'œufs
U-Net	20 min	0.377 (±0.090)	2.086 (±0.808)	357
YOLO26	15 min	0.401 (±0.073)	**1.910** (±1.020)	320
SegHVUNet	23 min	**0.414** (±0.092)	2.073 (±0.517)	306

Les résultats correspondent à la moyenne des 5-folds. L'écart type entre les folds est donné entre parenthèses. 326 œufs sont annotés sur les données de validation. La différence de surface correspond à la différence entre la surface médiane annotée et prédite de chaque échographie.

TABLE 3 – Résultats de la validation croisée sur les échographies longitudinales

Sur la première image de la figure 8, tous les modèles obtiennent un assez bon recouvrement des masques, en particulier SegHVUNet et U-Net avec un Dice respectivement de 0.73 et de 0.74 (contre 0.59 pour YOLO). Cependant, lorsque l'on s'intéresse aux œufs individuellement, on remarque que SegHVUNet et U-Net ont beaucoup plus de difficulté à les distinguer. Cela est dû, notamment, au fait qu'ils produisent initialement une segmentation sémantique, alors que YOLO réalise directement une segmentation d'instance. En effet, malgré le post-traitement de U-Net et les cartes H/V prédites de SegHVUNet, la séparation des œufs reste imparfaite. Ainsi, l'analyse des scores Dice est à regarder dans un contexte plus global, puisqu'un bon recouvrement ne garantit pas une bonne séparation des œufs. Il aurait fallu, pour cela, calculer un Dice individuel par œuf.
En revanche, lorsque l'on s'intéresse aux surfaces médianes des œufs, on retrouve une surface médiane supérieure pour YOLO (environ 9 mm^2), alors que les annotations et SegHVUNet présentent toutes deux une médiane proche de 7 mm^2. Cela montre une limite de l'utilisation de la médiane seule : des œufs dont la taille est surestimée et sous-estimée peuvent se compenser et donner une surface médiane similaire. Ce phénomène se retrouve aussi avec l'image de la deuxième ligne de la figure 8, qui montre que U-Net et SegHVUNet ont prédit les œufs au bon endroit et avec les bonnes délimitations, mais en regroupant plusieurs œufs. Malgré cela, U-Net obtient une surface médiane proche de celle des annotations. Cependant, malgré le fait que YOLO arrive bien à séparer les œufs, il a tendance à les surestimer et à souvent prédire uniquement les plus gros œufs. Cela induit au global une surestimation de la surface médiane estimée. A l'inverse, U-Net a tendance à sous-estimer la surface médiane. SegHVUNet présente, quant à lui, un comportement plus équilibré entre surestimation et sous-estimation. Les figures des différences de surfaces des modèles par poisson sont disponibles en annexe (figures 30, 31 et 32).

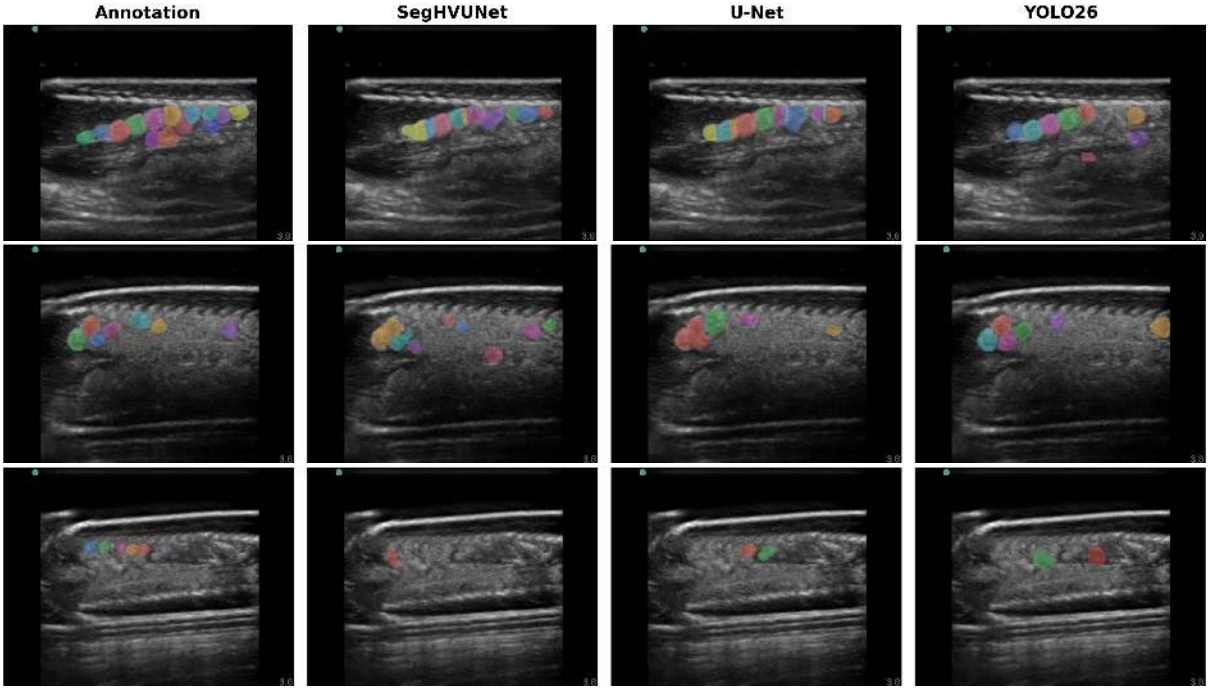


FIGURE 8 – Visualisation des prédictions des modèles sur les échographies longitudinales

Note : Les deux premières échographies montrent un assez bon recouvrement des œufs pour l'ensemble des modèles, mais une meilleure séparation des instances pour YOLO. Sur la troisième échographie, aucun modèle ne parvient à segmenter correctement les œufs de petite taille.

A cela s'ajoute également une autre limite, plus commune à tous les modèles. Lorsque les œufs sont petits, les modèles ont tendance à prédire de mauvais œufs, souvent plus gros que ceux annotés. La raison à cela est qu'un expert peut s'appuyer sur ses connaissances biologiques pour interpréter les échographies où les œufs sont peu visibles, comme des poissons avec de petits œufs. Les modèles ne disposent pas explicitement de cette information et peuvent donc identifier comme des œufs des structures qui y ressemblent, mais dont les dimensions sont incohérentes avec celles attendues. La dernière image de la figure 8 illustre ce problème.
Une deuxième limite est qu'un expert, lorsqu'il annote des œufs, sait qu'ils doivent avoir à peu près la même taille. Les modèles n'intègrent pas explicitement cette contrainte et peuvent ainsi prédire des œufs de tailles très différentes sur une même image. À noter que la variabilité entre les œufs d'une échographie existe, notamment lorsque seule une partie de l'œuf est présente.

Finalement, le modèle YOLO a été conservé pour estimer le volume des œufs des poissons. Pour cette tâche, le critère principal n'a pas été le recouvrement global de l'ensemble des masques, mais la capacité à obtenir des instances individuelles d'œufs dont le volume peut être calculé. Malgré sa tendance à surestimer leur volume, YOLO présente l'avantage de bien séparer les œufs entre eux. De plus, le fait qu'il prédise principalement les plus gros œufs pourrait être plus un avantage qu'un inconvénient pour l'estimation finale de la fécondité, si l'on considère que certains œufs apparaissent plus petits qu'ils ne le sont réellement.

4.5 Prédictions sur de nouvelles données

4.5.1 Performance des modèles retenus sur l'échantillon de test

Les modèles finaux gardés pour estimer les volumes sont U-Net et YOLO. Ce sont également ceux choisis pour la visualisation des prédictions et le calcul final de la fécondité dans l'interface.

L'échantillon de test se compose de 8 poissons, correspondant à 24 échographies transversales et 11 échographies longitudinales, avec au moins un poisson tiré pour chaque année. Les prédictions sur l'échantillon de test ont été obtenues en faisant la moyenne des probabilités prédites pour chaque fold d'entraînement pour les modèles sémantiques, afin d'avoir une prédiction plus robuste. Pour le modèle d'instance, un modèle final a été réentraîné sur l'ensemble des données d'entraînement et de validation. L'entraînement a été arrêté après un nombre d'epochs correspondant au nombre moyen d'epochs obtenu sur les différents folds de validation.

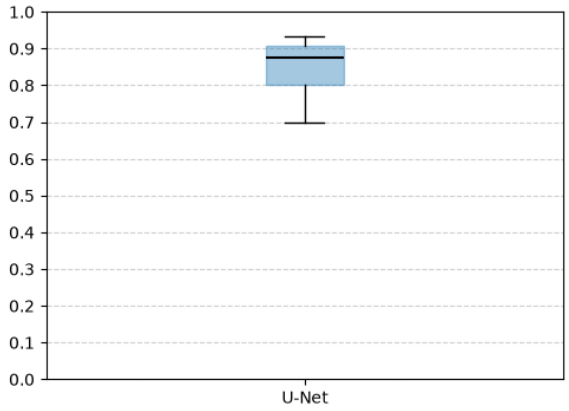


FIGURE 9 – Dice U-Net pour les 24 échographies transversales de test (gonades)

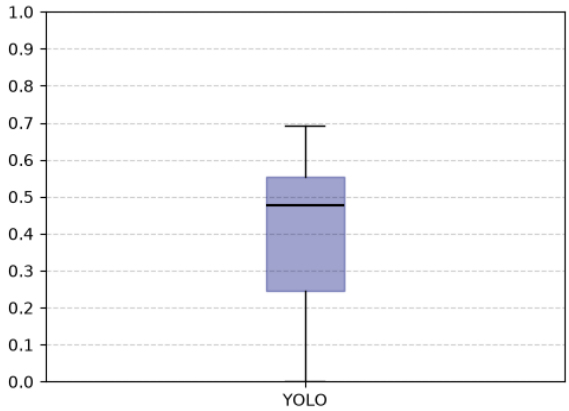


FIGURE 10 – Dice YOLO pour les 11 échographies longitudinales de test (œufs)

La figure 9 montre que les résultats obtenus pour les gonades sur l'échantillon de test sont globalement très bons. Contrairement aux données de validation, aucune image ne présente une segmentation très mauvaise, c'est-à-dire avec un score Dice inférieur à environ 0.6. Cela peut notamment s'expliquer par la composition de l'échantillon de test qui, de manière aléatoire, ne contient que des échographies sur lesquelles les gonades sont relativement bien visibles.
Les prédictions des œufs obtenues avec YOLO en figure 10 présentent également une distribution des scores Dice légèrement meilleure que sur l'échantillon de validation. On peut voir qu'il y a néanmoins des échographies avec un Dice égal à 0, donc sans aucun œuf retrouvé similaire aux annotations.

La figure 11 montre des exemples de prédictions des gonades et des œufs. La première image transversale montre le score Dice le plus faible obtenu sur les données de test, soit 0.68. À l'inverse, l'image située en dessous montre une des meilleures segmentations, avec un Dice de 0.91.
La première image longitudinale montre une situation où le score Dice obtenu est égal à 0. Cependant, bien que YOLO n'ait pas réussi à prédire les œufs annotés, l'œuf qu'il a prédit présente une surface cohérente avec celle des œufs annotés et pourrait correspondre à un œuf réel non annoté. Enfin, l'image située en dessous montre une échographie sur

laquelle le modèle a réussi à retrouver une majorité des œufs annotés avec un Dice de 0.67.

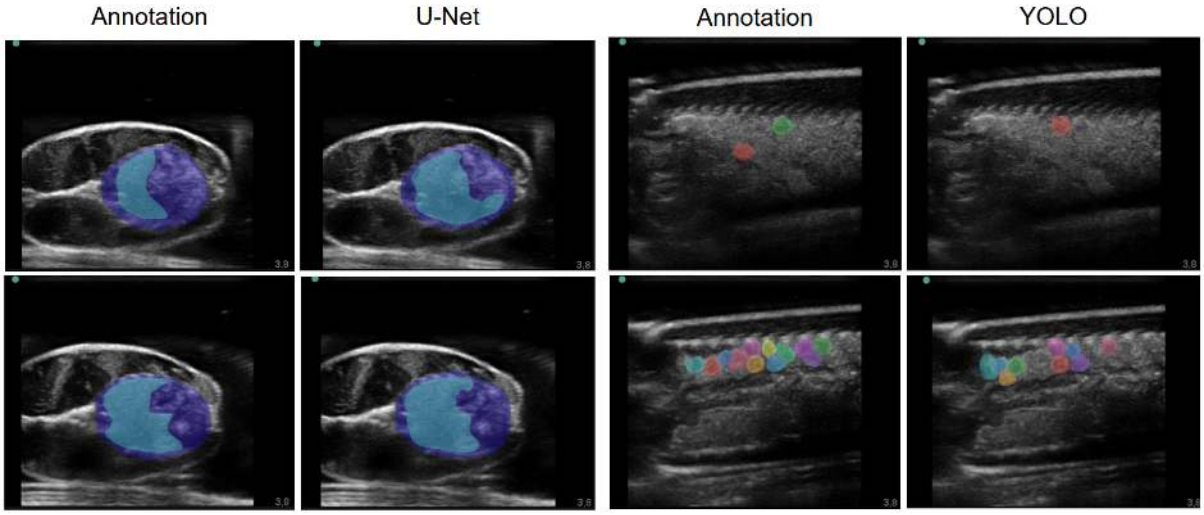


FIGURE 11 – Visualisation des prédictions U-Net et YOLO sur des échographies transversales et longitudinales

Les performances d'U-Net et de YOLO ne semblent donc pas se dégrader sur les données de test par rapport aux données de validation. Ces prédictions ont donc été utilisées pour estimer par la suite le volume des gonades, des œufs et la fécondité.

4.5.2 Estimation de la fécondité

Comme le montre la figure 12, les bonnes performances obtenues pour la segmentation des gonades sur l'échantillon de test se répercutent sur une bonne estimation de leur volume par poisson. On observe seulement une légère tendance à sous-estimer les volumes, mais avec des erreurs beaucoup plus faibles que sur l'échantillon de validation (voir figure 31 en annexe). Une erreur plus importante est toutefois observée pour le poisson 356434, qui présente une très grosse gonade et dont le volume a été davantage sous-estimé.

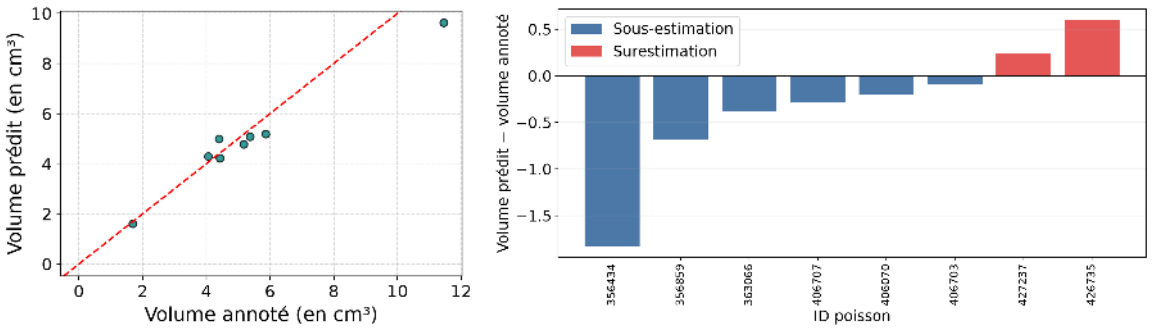
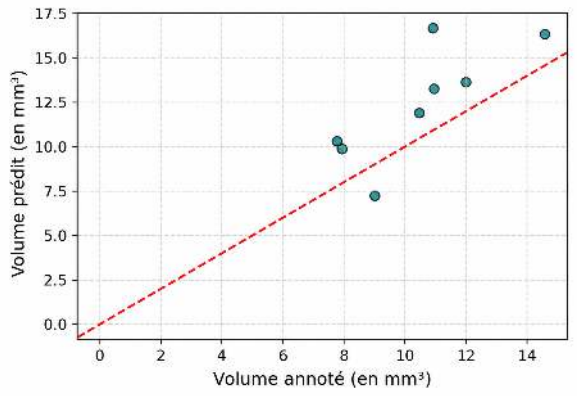


FIGURE 12 – Comparaison du volume des gonades estimé à partir des prédictions et des annotations

Données : les huit poissons du jeu de test.

Pour les œufs, on retrouve la tendance à la surestimation de YOLO déjà observée sur l'échantillon de validation (figure 13). Le poisson 356434 se distingue également par un volume moyen des œufs fortement surestimé. Il s'agit d'un poisson pour lequel l'emplacement des œufs a bien été retrouvé, mais leurs contours ont été élargis.



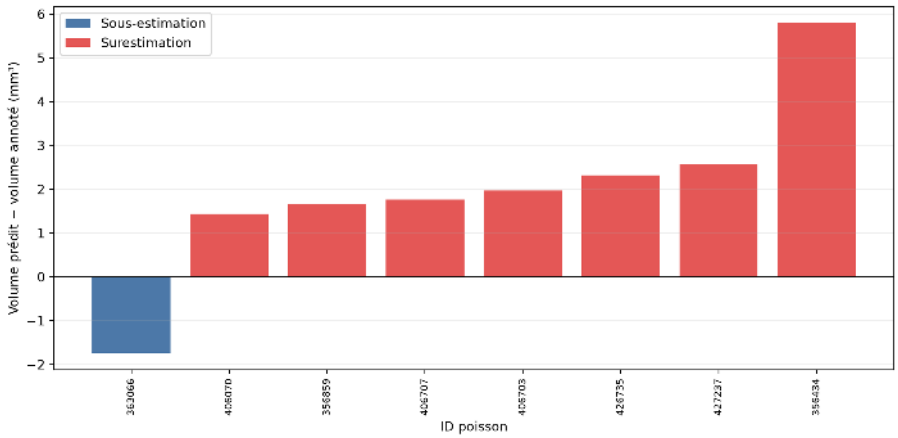


FIGURE 13 – Comparaison du volume médian des œufs estimé à partir des prédictions et des annotations

Données : les huit poissons du jeu de test.

Concernant l'estimation de la fécondité, les résultats sont relativement proches des valeurs obtenues à partir des annotations pour la majorité des poissons, avec une tendance générale à la sous-estimation, directement liée à la surestimation du volume médian des œufs (figure 14). On retrouve également le poisson 356434, pour lequel le volume des œufs a été fortement surestimé tandis que le volume des gonades a, à l'inverse, été sous-estimé. Ces deux erreurs se cumulent dans le calcul de la fécondité et conduisent à une estimation particulièrement éloignée de celle obtenue à partir des annotations.

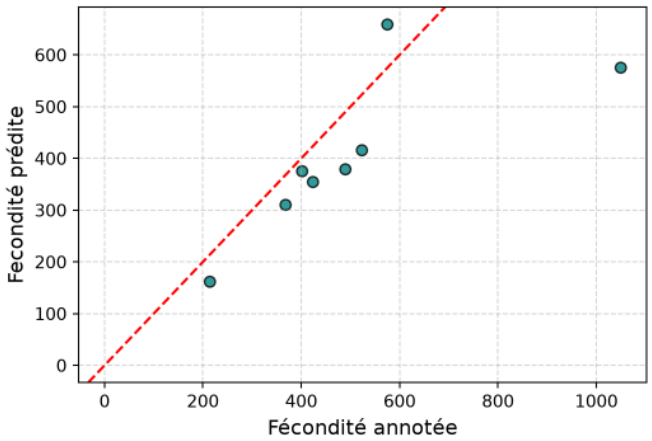


FIGURE 14 – Relation entre la fécondité prédite et la fécondité calculée à partir des annotations

Données : les huit poissons du jeu de test.

4.5.3 Vérification et utilisation des prédictions par les experts

L'interface graphique mise en place permet de faciliter l'utilisation des modèles par les experts. Elle permet de visualiser et de corriger les prédictions des modèles de segmentation, ainsi que d'obtenir directement les estimations des volumes et de la fécondité.

L'interface (figures 34 et 35 en annexe) permet d'importer de nouvelles échographies et de les segmenter automatiquement à l'aide des modèles entraînés sélectionnés : U-Net et YOLO. Elle renvoie ainsi, pour chaque poisson, une visualisation des prédictions obtenues sur chaque échographie, ainsi que les caractéristiques du poisson. L'utilisateur peut ensuite choisir d'exporter directement dans un fichier csv les volumes et la fécondité estimés à partir de ces prédictions.
Si les prédictions ne lui semblent pas suffisamment satisfaisantes, il peut les modifier directement dans l'interface à l'aide d'un outil de labellisation. Les annotations corrigées sont alors utilisées pour obtenir une nouvelle estimation de la fécondité.
L'utilisateur peut également évaluer manuellement la qualité de la segmentation des images. S'il considère qu'une prédiction n'est pas satisfaisante, il peut, en plus de la modifier, simplement la retirer du calcul des volumes. A l'inverse, les échographies dont la segmentation est considérée comme très bonne peuvent être ajoutées à la base d'apprentissage des modèles et être utilisées pour de futurs entraînements.

Puisqu'il n'est pas possible, sur de nouvelles données, de vérifier les résultats, un processus de détection d'anomalies a également été mis en place afin d'alerter l'utilisateur sur de potentielles erreurs. Ces alertes ne signifient pas nécessairement qu'une échographie est mal prédite, mais signalent que ce sont des échographies qu'il faudrait peut-être vérifier en priorité avant d'estimer la fécondité.
La première alerte, sur les données longitudinales, est déclenchée si le volume des œufs prédits est anormalement faible ou élevé par rapport à l'ensemble des prédictions. La deuxième alerte, sur les données transversales, consiste à vérifier si l'évolution de la surface des gonades au sein d'un même poisson est cohérente. En effet, comme montré sur la figure 36 en annexe, la surface des gonades est en général plus petite au niveau de l'opercule, puis augmente à Op+2 avant de rétrécir progressivement jusqu'à la fin de la gonade. A noter que c'est le schéma classique, mais qu'il peut y avoir des exceptions, comme c'est le cas des trois poissons visibles sur la figure. En effet, sur les échographies Op+2 de ces poissons, présentées en annexe sur la figure 37, une grande zone a été annotée comme "Ignore gonade", car l'experte n'a pas pu déterminer avec certitude si une gonade était présente. Or, au vu de l'évolution des surfaces observée sur les autres coupes du même poisson, on s'attendrait à retrouver une gonade plus importante à cette position. Cette analyse a posteriori suggère donc qu'une gonade est probablement présente dans cette zone, et pourrait permettre de revoir l'annotation initiale.

5 Discussion

Les résultats obtenus permettent une première estimation de la fécondité des truites de la rivière de l'Oir. Cette estimation est cependant limitée par de nombreuses incertitudes tout au long du processus. La première limite est la qualité des données elles-mêmes, puis les erreurs d'annotation et de calcul du volume, et enfin les erreurs dues aux prédictions des modèles. La vérité terrain ne pouvant être connue, l'ampleur de ces incertitudes reste inconnue.

Erreur des données et incertitude des annotations
Les premières limites sont liées à l'acquisition des données. Les échographies sont prises sur le terrain, dans un temps limité, ce qui fait que les mesures ne sont pas toujours exactes ni précises. Les échographies acquises peuvent également être difficilement lisibles, notamment avec la présence d'artefacts qui masquent une partie de l'image. Ces limites de qualité d'image rendent les contours des structures ambigus et compliquent leur annotation par les experts. Les annotations utilisées comme référence ne sont donc pas toujours exactes ni précises au pixel près. L'ajout de la classe "Ignore gonade" illustre cette incertitude.
De plus, l'annotation manuelle des échographies étant longue et complexe, relativement peu de données ont pu être annotées pour chaque type d'échographie, ce qui peut restreindre l'apprentissage des modèles et leur capacité à généraliser à des cas peu représentés.

Robustesse de la méthode de calcul de la fécondité
La méthode utilisée pour estimer la fécondité peut également être questionnée. Elle permet une première estimation de la fécondité, jusqu'alors inconnue, mais elle se base pour cela sur plusieurs hypothèses.
Une de ces hypothèses repose sur l'approximation des volumes par des formes géométriques simples. Ces approximations engendrent inévitablement des erreurs dans l'estimation du volume. Le faible nombre de coupes par poisson (entre 2 et 7) limite également la précision de cette estimation.
Nous avons également supposé que les œufs d'un même poisson faisaient la même taille, sauf qu'en pratique, on observe des tailles différentes (figure 38 en annexe). Cela s'explique à la fois par les coupes qui ne passent pas par le diamètre maximal des œufs (problème de stéréologie de Wicksell [14]), par le fait que les œufs se trouvent à des distances différentes de l'échographe et par le fait que certains œufs sont en partie masqués. Ainsi, comme tous ces problèmes auraient tendance à sous-estimer la taille réelle des œufs, un choix pertinent aurait également été de prendre la moyenne ou la médiane des plus gros œufs uniquement, plutôt que la médiane de l'ensemble des œufs. Par exemple, dans l'article de Yari et al. de 2024 [6] sur la maturation des œufs, les auteurs ont choisi de ne conserver que les 20% plus gros œufs afin d'éviter le biais dû aux coupes non perpendiculaires du poisson.

Pour mieux quantifier une partie de l'incertitude autour de l'estimation de la fécondité, une approche par bootstrap aurait également pu être mise en place. Elle aurait consisté à rééchantillonner les œufs observés pour un même poisson afin d'obtenir une distribution du volume représentatif d'un œuf, puis un intervalle d'incertitude autour de la fécondité estimée. Cet intervalle aurait ainsi permis de caractériser l'incertitude liée à l'échantillonnage des œufs observés. Cette analyse n'a toutefois pas été réalisée par manque de temps.

Enfin, l'intégralité du volume de la gonade n'est peut-être pas occupée par les œufs. Ceux-ci peuvent notamment être moins denses au début et à la fin de la gonade. L'estimation de la fécondité à partir du simple rapport entre le volume des gonades et le volume représentatif d'un œuf constitue donc également une approximation.

Limites des prédictions des modèles
Concernant les modèles, des résultats plus robustes auraient pu être obtenus en réalisant plusieurs répétitions de chaque fold avec des initialisations différentes. Cela aurait permis de mieux caractériser la variabilité des performances due à l'entrainement et de mieux distinguer les écarts réellement dus aux architectures. Cela aurait cependant fortement augmenté le temps de calcul nécessaire.

Le résultat le plus contre-intuitif est celui de CSA-Net sur l'échantillon de validation. Les faibles performances de ce modèle, comparées à celles des autres modèles, peuvent s'expliquer par les caractéristiques des données elles-mêmes. En effet, les coupes successives d'un même poisson sont relativement éloignées les unes des autres (environ 2 cm). Ainsi, la gonade change énormément d'une coupe à l'autre, ce qui limite la continuité anatomique observable entre deux images. De plus, la qualité des échographies peut varier fortement d'une coupe à l'autre, par exemple avec la présence d'artefacts sur une image, mais pas sur la suivante. Une solution pour bénéficier de ces informations sans perturber la segmentation serait d'ajuster le protocole d'acquisition des données pour les prochaines années pour prendre les échographies les plus resserrées possible, tous les 1 cm par exemple et regarder si les prédictions s'améliorent.
Cependant, à l'exception de CSA-Net, les modèles appliqués aux échographies transversales ont obtenu des résultats très similaires. Cette faible différence entre les architectures peut laisser penser que la qualité et la quantité des données annotées constitueraient un facteur limitant plus important, quant à l'amélioration des performances, que le choix précis de l'architecture. Un point d'amélioration des résultats serait donc plus sur un travail approfondi sur l'annotation et la préparation des données.

Enfin, l'évaluation des modèles de segmentation des œufs aurait pu être complétée par des métriques plus adaptées à la segmentation d'instances. Les scores Dice utilisés évaluent principalement le recouvrement global des masques et ne permettent pas de mesurer directement la capacité des modèles à séparer correctement les différents œufs. Il aurait par exemple été possible d'apparier chaque œuf prédit avec un œuf annoté, puis de calculer un Dice au niveau de chaque instance. Cette approche présente toutefois une limite dans notre cas, puisque l'objectif n'était pas nécessairement de segmenter l'ensemble des œufs présents. De plus, tous les œufs visibles n'étaient pas systématiquement annotés et l'estimation de la fécondité ne nécessite d'obtenir que quelques œufs correctement segmentés pour estimer un volume représentatif.

Vérification de la fécondité dans la littérature
Puisque la vérité terrain n'est pas connue, la justesse des résultats ne peut pas être confirmée. Il faut donc utiliser d'autres moyens pour évaluer au mieux la cohérence des résultats obtenus. Une des possibilités est de se baser sur les résultats et relations disponibles dans la littérature. Cependant, l'une des principales limites de ces comparaisons est que les populations étudiées ne sont généralement pas les mêmes, alors que la fécondité peut fortement varier d'une population à l'autre. En effet, les relations observées peuvent être très différentes selon les études et les conditions environnementales.

Il y a cependant un consensus sur le fait que la fécondité des truites est généralement liée à la taille de la femelle. La longueur de la truite et sa fécondité sont ainsi généralement corrélées positivement. Cependant, lorsque l'on visualise cette relation sur l'ensemble des données annotées, on observe une faible corrélation linéaire à 0.46.

De plus, une étude a déjà été réalisée par Acolas et al. en 2008 [15] sur les truites communes de l'Oir. Sur les 46 femelles résidentes analysées, ils ont établi une relation entre la longueur et la fécondité des truites ($R^2 = 0{,}70$) : $\mathrm{F} = 4.0867 * \mathrm{Lf(mm)} - 522.26$.
Lorsque l'on applique cette relation aux données annotées, les fécondités estimées sont souvent supérieures aux valeurs prédites par la relation d'Acolas et al. (figure 39 en annexe). Cette tendance à la surestimation peut provenir d'une sous-estimation du volume moyen des œufs, d'une surestimation du volume des gonades ou d'une combinaison de ces deux erreurs. Or, au regard des discussions précédentes sur le fait que l'estimation du volume d'un œuf pourrait être améliorée en privilégiant les œufs les plus gros, une sous-estimation du volume des œufs pourrait en grande partie contribuer à cet écart. Il serait ainsi intéressant pour la suite d'évaluer la fécondité à partir du volume médian des plus gros œufs pour voir si on obtiendrait des résultats plus proches de la relation d'Acolas. On peut également remarquer deux individus qui se distinguent des autres par une fécondité bien supérieure à celle calculée à partir de leur longueur. Ces estimations ne semblent pas non plus suivre les fécondités estimées par notre méthode et peuvent très probablement correspondre à des erreurs d'annotation.

Cette comparaison doit toutefois être interprétée avec prudence. Bien que l'étude ait été réalisée sur l'Oir, les conditions sont différentes. Par exemple, les truites ont été pêchées entre novembre et janvier, contre octobre pour nos truites. De plus, les truites résidentes analysées sont plus grandes que celles de notre échantillon : entre 16 et 40 cm contre 15 à 32 cm pour celles des échographies. Enfin, cette étude date de 2008, or les conditions de la rivière ont pu fortement changer en plus de 10 ans.

6 Conclusion

L'objectif du stage était de mettre en place un outil permettant d'aider les experts à annoter les échographies de truites, ainsi que de développer un pipeline permettant d'estimer leur fécondité. La problématique principale du stage consistait ainsi à appliquer des modèles de deep learning pour segmenter les gonades et les œufs présents sur les échographies.
Les échographies étant des images difficiles à segmenter, plusieurs modèles ont été comparés afin d'obtenir des segmentations les plus fidèles possible aux annotations des experts. Différentes architectures et différents post-traitements ont ainsi été appliqués et testés pour répondre aux différentes problématiques liées aux données : spécificités des images échographiques, coupes séquentielles, segmentation individuelle des œufs et contraintes morphologiques des structures.
Finalement, les modèles les plus simples pour chaque tâche de segmentation ont été retenus : U-Net pour la segmentation des gonades et YOLO pour celle des œufs. Les modèles permettent d'obtenir des segmentations globalement satisfaisantes, avec cependant des performances très mauvaises sur certaines échographies particulièrement difficiles. YOLO présente également une tendance générale à surestimer le volume des œufs. Malgré ces limites, les modèles entraînés permettent de fournir des prédictions sur de nouvelles données et ainsi proposer aux experts une première segmentation, qu'ils peuvent ensuite vérifier et corriger grâce à l'interface proposée.
Les prédictions permettent également d'obtenir une première estimation de la fécondité des truites étudiées. Ces estimations sont toutefois à interpréter avec prudence car elles semblent élevées et ne semblent pas suivre la relation fécondité-longueur attendue. Le travail réalisé permet néanmoins de proposer une première chaîne de traitement complète, allant de l'échographie non traitée jusqu'à l'estimation du nombre d'œufs.

Concernant les perspectives, une première amélioration possible serait d'effectuer un travail de vérification et de correction des annotations afin d'obtenir une base d'apprentissage aussi fiable et précise que possible. Les résultats obtenus au cours du stage suggèrent en effet que la qualité et la quantité des données annotées constituent probablement un facteur important pour l'amélioration des performances des modèles.
Il serait également intéressant d'exploiter davantage les nombreuses échographies non annotées à l'aide de méthodes auto-supervisées ou semi-supervisées, comme le pré-entraînement masqué proposé par He et al. [16]. Ces approches n'ont pas été étudiées au cours du stage par manque de temps, mais constituent une piste intéressante pour améliorer les performances sans nécessiter l'annotation manuelle d'un grand nombre d'images supplémentaires.
La méthode d'estimation de la fécondité pourrait également être ajustée et optimisée pour qu'elle corresponde mieux à la réalité biologique. D'un point de vue écologique, il serait également intéressant d'étudier l'évolution de la fécondité au cours des différentes années afin d'identifier d'éventuelles variations temporelles.

Enfin, une perspective particulièrement intéressante serait d'appliquer le pipeline mis en place à des truites issues de pisciculture. Ces poissons ont deux avantages majeurs. D'une part, ils sont plus gros, avec probablement des organes et des œufs plus visibles et donc des structures plus nettes. D'autre part, les individus sont, à terme, destinés à être sacrifiés, ce qui permettrait de mesurer directement le nombre réel d'œufs. Cela permettrait d'obtenir une vérité terrain pour évaluer la segmentation mais également l'estimation finale de la fécondité.

Annexe

A - Présentation des données

A.1 - Acquisition des données



FIGURE 15 – Processus d'acquisition des échographies des truites sur le terrain

Note : La truite est positionnée dans un bac contenant de l'eau avec de l'anesthésiant pour la réalisation de l'échographie. On peut également voir à l'avant du bac le mètre ruban qui peut s'attacher à la sonde pour mesurer la longueur de la gonade et repérer la position de la coupe.

A.2 - Exemple d'artefacts



FIGURE 16 – Exemple d'une échographie avec plusieurs artefacts

Note : On observe plusieurs artefacts blancs qui masquent des parties de l'échographie. Comme l'échographie est prise au début du poisson, ces artefacts proviennent probablement de l'opercule.

A.3 - Anatomie du poisson

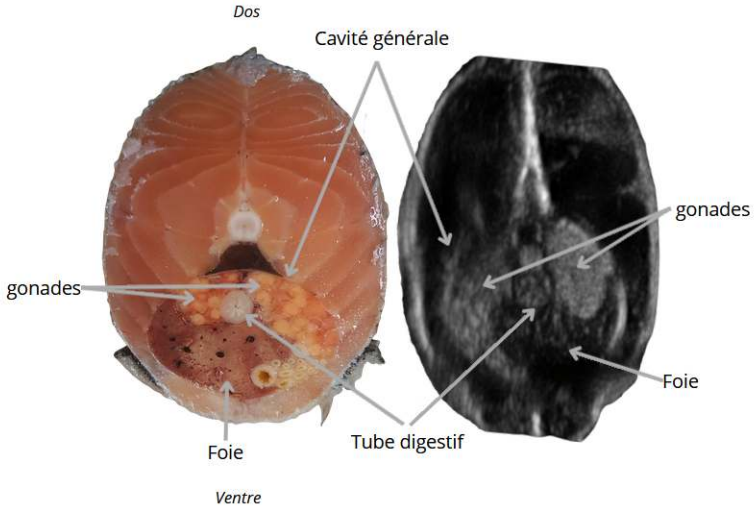


FIGURE 17 – Schéma anatomique de la cavité d'une truite femelle : observation directe vs échographie

Note : Les coupes ont été faites au début du poisson, où les gonades partagent la cavité avec d'autres organes, notamment le foie et le tube digestif.

B - Présentation des modèles

B.1 - YOLO

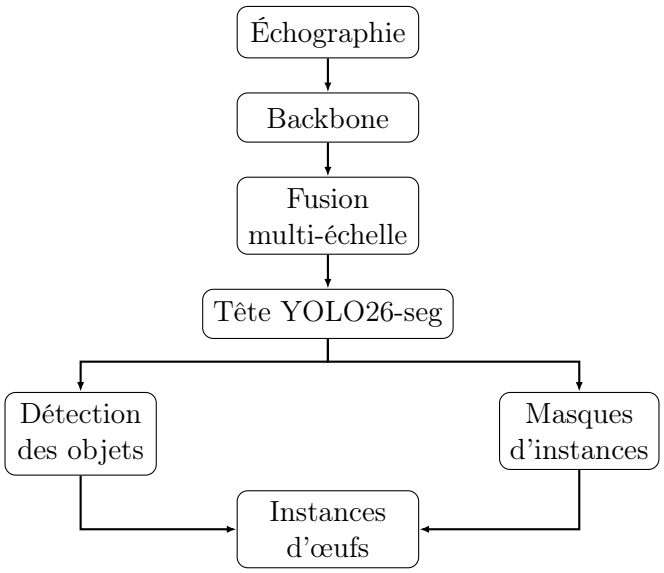


FIGURE 18 – Schéma simplifié de l'architecture de YOLO26-seg.

YOLO26-seg est un modèle de segmentation d'instances. L'image est traitée par un backbone convolutionnel afin d'en extraire progressivement des caractéristiques à différentes résolutions, qui sont ensuite fusionnées. Cela permet au modèle de combiner des informations spatiales fines avec des informations plus globales et ainsi de prendre en compte des objets de tailles différentes. Les caractéristiques obtenues sont ensuite transmises à la tête YOLO26-seg. A la différence des modèles de segmentation sémantique, celle-ci réalise à la fois la détection des objets, qui permet de les localiser et de les classifier, et la prédiction

de leurs masques. L'association de ces deux prédictions permet ainsi d'obtenir un masque distinct pour chaque œuf détecté. Le principe général du modèle est schématisé dans la figure 18.

B.2 - U-Net

U-Net est un réseau de neurones convolutionnel qui repose sur une architecture encodeur-décodeur en forme de U. Chaque niveau de l'architecture est composé de plusieurs couches de convolution permettant d'extraire progressivement des caractéristiques de l'image. L'encodeur réduit progressivement la résolution spatiale de l'image afin d'en extraire des caractéristiques de plus haut niveau, tandis que le décodeur reconstruit progressivement la résolution pour produire le masque de segmentation. Des connexions de saut (*skip connections*) relient les différents niveaux de l'encodeur aux niveaux correspondants du décodeur, ce qui permet de conserver les informations spatiales fines nécessaires à la segmentation. Une architecture U-Net classique est présentée en figure 19.

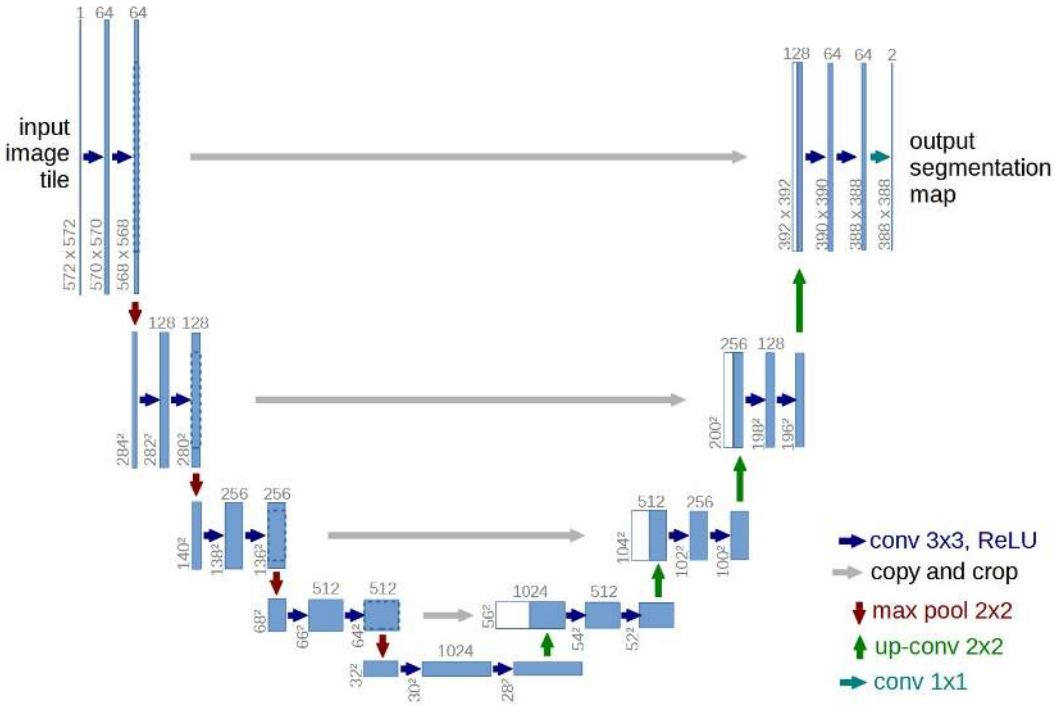


FIGURE 19 – Architecture U-Net classique, d'après Ronneberger et al. [3]

B.3 - ResAttentionUNet++

L'encodeur ResNet-34 est pré-entraîné sur ImageNet et extrait progressivement des cartes de caractéristiques à différentes résolutions. Les blocs résiduels, issus de ResNet, facilitent notamment la propagation du gradient dans un réseau relativement profond. Le décodeur U-Net++ a quant à lui l'avantage d'utiliser des connexions de sauts imbriquées et denses, plutôt qu'une seule connexion directe entre les mêmes niveaux de l'encodeur et du décodeur, comme pour U-Net. Cela permet de combiner plus progressivement les caractéristiques provenant des différentes profondeurs du réseau. Le mécanisme d'attention par canal permet, après extraction des caractéristiques, de déterminer lesquelles sont les plus importantes. Chaque canal reçoit un poids : les caractéristiques jugées utiles à la segmentation sont ainsi renforcées, tandis que les autres sont atténuées. Enfin, le mécanisme d'attention spatiale vise à déterminer où se trouve l'information importante dans l'image, en renforçant ou atténuant certaines positions de l'image en conséquence.

B.4 - CSA-Net

CSA-Net est un modèle de segmentation "2.5D" qui utilise l'information des coupes consécutives pour segmenter la coupe centrale. Il se base sur des mécanismes d'attention pour extraire les relations importantes entre et au sein des coupes.
Le principe est que, pour chaque ensemble de trois coupes consécutives, un même CNN encodeur (ResNet-50) est appliqué à chacune des coupes afin d'en extraire les caractéristiques. Un mécanisme de *Cross-Slice Attention* est par la suite appliqué entre la coupe centrale et chacune de ses voisines afin d'identifier des correspondances entre les régions de différentes coupes. L'attention est calculée entre chaque position spatiale de la coupe voisine et chaque position de la coupe centrale. La *Cross-Slice Attention* est multi-tête (16 dans l'article), ce qui permet à chaque tête d'apprendre des types de relations différents. Le modèle applique également un mécanisme d'*In-slice attention* pour apprendre les relations entre les différentes régions de la coupe centrale via un mécanisme de self-attention classique. Les trois sorties d'attention sont ensuite fusionnées puis transmises à un Vision Transformer pré-entraîné, qui permet d'exploiter les informations globales obtenues. Enfin, un décodeur reconstruit progressivement la résolution spatiale afin de produire la segmentation de la coupe centrale. Les étapes sont simplifiées en figure 20.

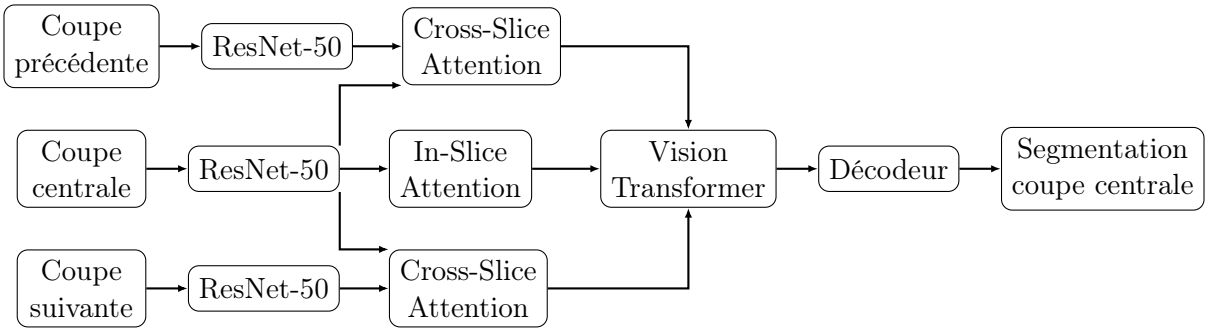


FIGURE 20 – Schéma simplifié de l'architecture CSA-Net.

B.5 - SegHVUNet

Le modèle proposé par Yari et al. [6] repose sur une architecture U-Net classique qui a été améliorée afin de mieux segmenter et séparer les œufs de saumon atlantique. La particularité du modèle est qu'en plus de prédire un masque binaire des œufs, il prédit également deux cartes horizontales et verticales (H/V), qui codent, pour chaque pixel, sa position par rapport au centre de l'œuf auquel il appartient. Les valeurs de ces cartes sont normalisées entre -1 et 1, le centre de chaque œuf étant 0.
Une fois prédites, ces cartes sont utilisées lors d'une étape de post-traitement afin de séparer les œufs adjacents. Pour cela, des filtres de Sobel sont appliqués aux cartes H/V afin de détecter les zones où les valeurs changent brutalement. Ces zones sont alors interprétées comme des frontières entre les instances. Pour obtenir une seule carte de frontières, le maximum est ensuite pris, pixel par pixel, entre les réponses obtenues à partir des cartes H et V. Elle est ensuite combinée avec le masque de segmentation binaire pour obtenir les instances individuelles d'œufs. Une tête de classification est parallèlement ajoutée à l'encodeur pour prédire le stade de maturation des œufs. Cette partie n'a pas été reproduite dans notre cas, puisque seule la segmentation des œufs nous intéresse. Le principe général du modèle est schématisé dans la figure 21.

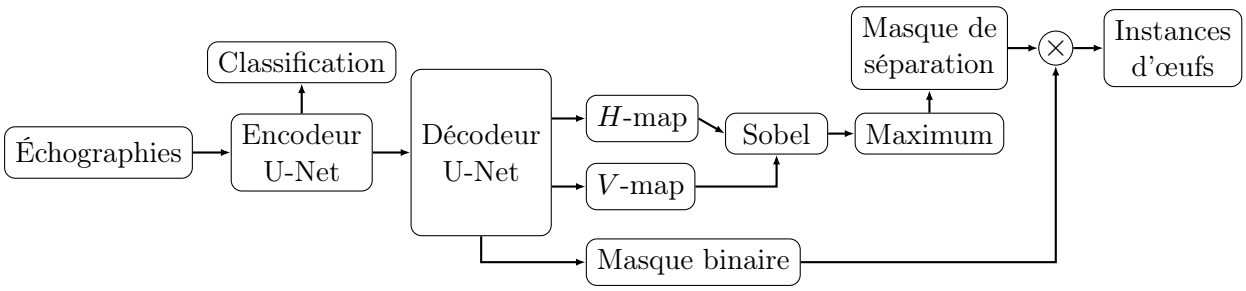


FIGURE 21 – Principe général du modèle SegHVUNet

B.6 - OpenUS

OpenUS est un modèle pré-entraîné reposant sur une architecture Vision Mamba, capable de capturer à la fois des informations locales et des dépendances spatiales à longue distance. L'image est d'abord découpée en patches, puis traitée à plusieurs résolutions grâce à des étapes successives de sous-échantillonnage, ce qui permet au modèle d'extraire des caractéristiques à différentes échelles. Vision Mamba est une architecture issue des *State Space Models (SSM)*, qui permettent de faire circuler efficacement l'information sur de grandes distances dans l'image et ainsi de mieux prendre en compte le contexte spatial global.
OpenUS est pré-entraîné directement sur des échographies via un apprentissage auto-supervisé, c'est-à-dire qu'il apprend à extraire des caractéristiques à partir des images sans avoir besoin de masques annotés. Il utilise pour cela un système enseignant-étudiant. Les paramètres de l'enseignant sont progressivement mis à jour à partir de ceux de l'étudiant par moyenne mobile exponentielle (EMA), ce qui permet d'obtenir des représentations de référence plus stables. L'apprentissage repose sur deux processus. Le premier est l'apprentissage contrastif (*contrastive learning*), qui consiste à apprendre des représentations similaires pour différentes vues d'une même image. Cela permet au modèle d'apprendre des structures globales. L'enseignant fournit des représentations de référence, et l'étudiant apprend à s'en rapprocher à partir de vues différentes de la même image.
Le second processus est la reconstruction d'une image masquée (*Masked Image Modeling*), qui consiste à cacher certaines régions de l'image puis à demander au modèle de les reconstruire. Cela favorise l'apprentissage de caractéristiques plus fines au niveau des pixels. L'élève reçoit ainsi des images partiellement masquées et utilise les régions visibles pour reconstruire les régions cachées. Les auteurs proposent également un masquage auto-adaptatif (*Self-Adaptive Masking*). Contrairement à un masquage entièrement aléatoire, les régions à masquer sont progressivement sélectionnées en fonction de leur importance et de leur difficulté. Pour cela, l'enseignant identifie les régions importantes grâce à une carte d'attention, tandis que l'erreur de reconstruction de l'élève permet d'identifier les régions difficiles à reconstruire. Ces deux informations sont combinées pour sélectionner progressivement les régions les plus pertinentes à masquer. Un schéma simplifié du pré-entraînement est proposé en figure 22 afin de mieux visualiser les différents processus et les liens entre eux.

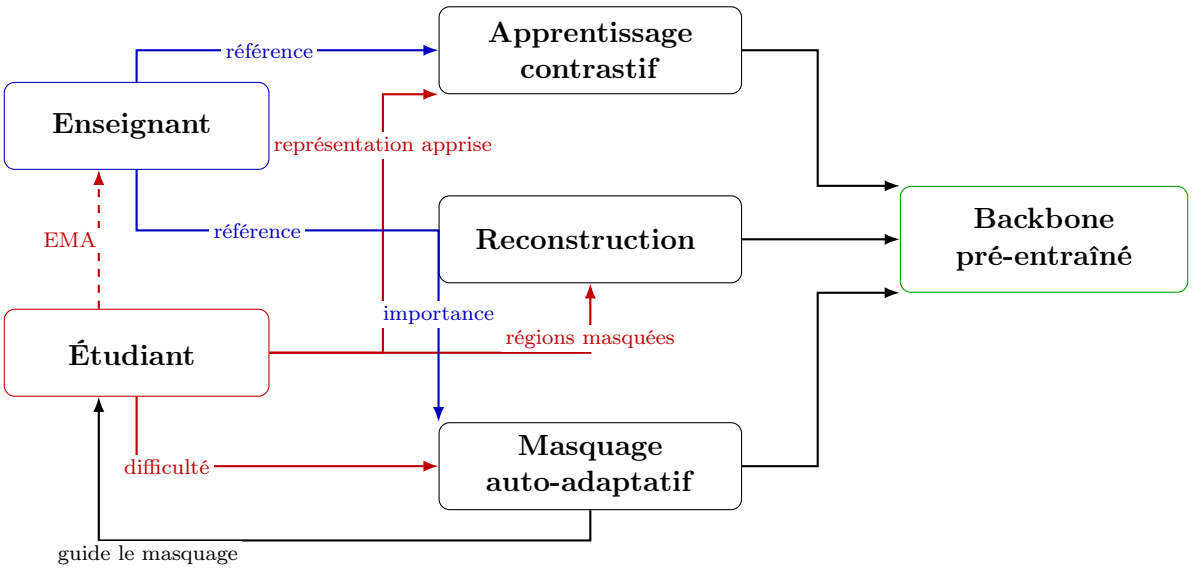


FIGURE 22 – Schéma simplifié du pré-entraînement d'OpenUS.

C - Post-traitement

C.1 - Ouverture et fermeture

La morphologie mathématique est utilisée pour le traitement d'images. Les deux opérations de base sont la dilatation et l'érosion. L'idée est de traiter un ensemble à partir d'un autre ensemble, appelé *élément structurant*. Dans notre cas, l'ensemble structurant choisi est une ellipse de taille différente selon les classes : 5x5 pixels pour la cavité, 13x13 pour la gonade et 3x3 pour les œufs.
Soit X un ensemble représentant les pixels appartenant à un objet et B un élément structurant. La dilatation morphologique de X avec B est définie par :

$$X \oplus B = \{x + b \mid x \in X,\ b \in B\}$$

La dilatation agrandit les objets du masque et permet de combler de petits espaces entre eux.
L'érosion est définie par :

$$X \ominus B = \{x \mid B_x \subseteq X\},$$

où B_x représente l'élément structurant B placé au point x. L'érosion réduit les objets du masque et permet de supprimer de petites structures.

Il est possible de combiner ces deux opérateurs pour obtenir des opérateurs plus complexes. La fermeture correspond à une dilatation suivie d'une érosion :

$$X \bullet B = (X \oplus B) \ominus B.$$

Elle permet de combler de petits trous ou discontinuités et de relier des régions proches. L'ouverture correspond à une érosion suivie d'une dilatation :

$$X \circ B = (X \ominus B) \oplus B.$$

Elle permet de supprimer de petites irrégularités ou composantes tout en conservant globalement la forme des objets.

C.2 - Méthode de séparation des œufs

Une fois les prédictions des œufs obtenues, il faut séparer les œufs qui se touchent pour obtenir des instances individuelles (figures 23 et 24). La première étape consiste à calculer, pour chaque pixel appartenant à la classe "œufs", sa distance au fond de l'image. Les pixels au centre des œufs sont ainsi ceux avec les distances les plus élevées, et inversement pour ceux proches des contours. A partir de ces distances, on peut obtenir des maxima locaux comme marqueurs des centres des œufs.
Une méthode de ligne de partage des eaux (*watershed*) est ensuite appliquée. La carte de distance de chaque pixel au fond peut être interprétée comme un relief dans lequel les valeurs élevées correspondent à des sommets. En inversant cette carte, les centres des œufs deviennent des minima et on obtient des vallées. Le principe du watershed consiste alors à simuler une inondation à partir de ces minima : le niveau de l'eau monte progressivement et, lorsque deux bassins de marqueurs différents se rencontrent, une frontière est créée entre eux (figure 25). Les bassins versants finaux obtenus correspondent aux régions des partitions. Dans le cas des œufs, ces frontières se placent généralement au niveau des zones de contact ou de rétrécissement entre plusieurs œufs.

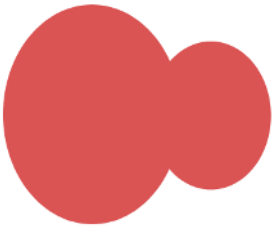


FIGURE 23 – Exemple de deux œufs fusionnés

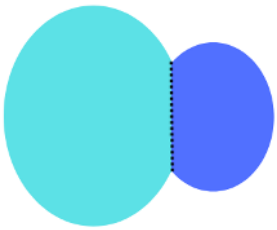


FIGURE 24 – Séparation attendue des deux œufs

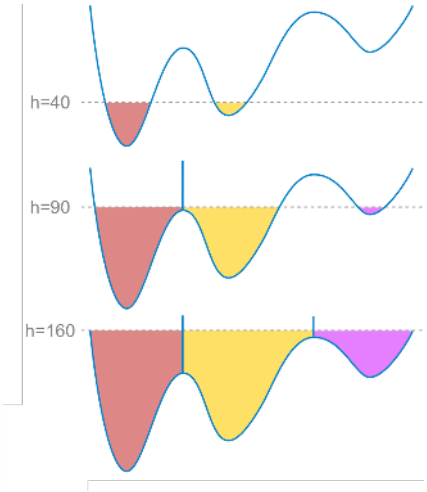


FIGURE 25 – Principe de la méthode de ligne de partage des eaux

Note : À partir des minima du relief, le niveau d'eau augmente progressivement. Lorsque deux bassins se rencontrent, une ligne de partage des eaux est créée afin de séparer les différentes régions.

En pratique, il est possible de fixer une distance minimale entre deux maxima locaux utilisés comme marqueurs. Une valeur trop faible peut séparer un même œuf en plusieurs instances, tandis qu'une valeur trop élevée peut fusionner des œufs voisins. La valeur retenue pour les œufs est de 6 pixels.

C.3 - Exemples du post-traitement

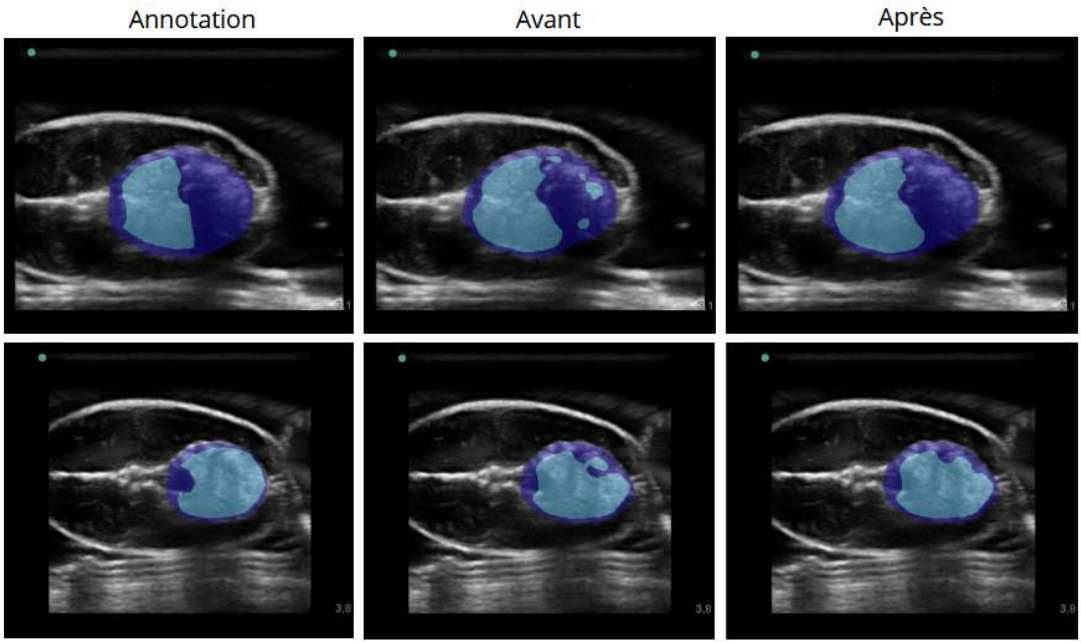


FIGURE 26 – Exemple de prédictions avant et après application du post-traitement sur des échographies longitudinales

Note : Sur la première échographie, les îlots de gonade isolés n'ont pas été conservés. Sur la seconde, une excroissance de la gonade a été fusionnée avec la composante principale.

D - Préparation des données et choix des hyperparamètres

D.1 - Choix des classes

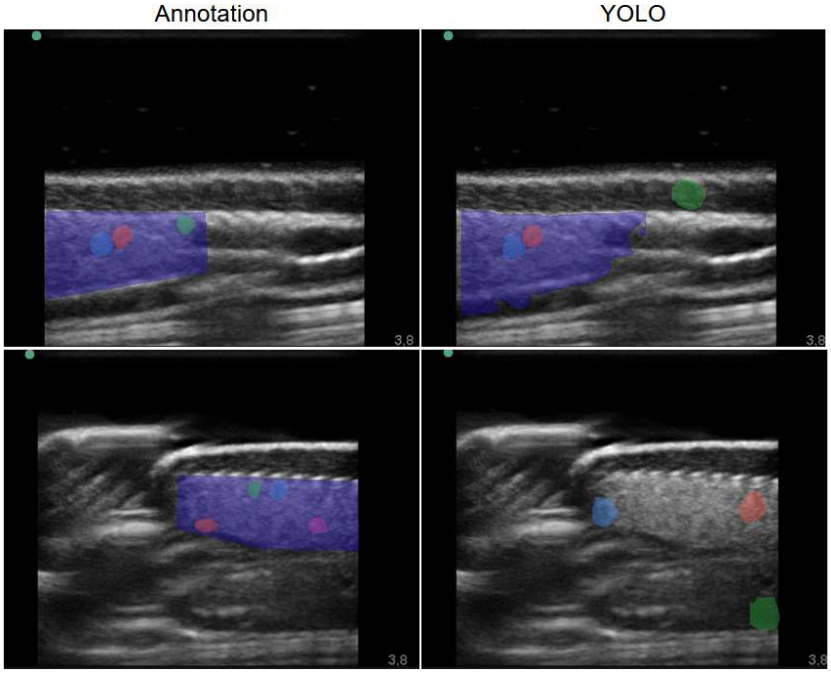


FIGURE 27 – Visualisation des prédictions YOLO26 avec l'ajout de la classe gonade

Note : La première image montre que l'ajout de la classe gonade ne permet pas toujours d'empêcher la prédiction d'œufs en dehors de la gonade. La deuxième image montre que la gonade elle-même n'est par ailleurs pas toujours prédite.

D.2 - Application du Conditional Metadata Embedding

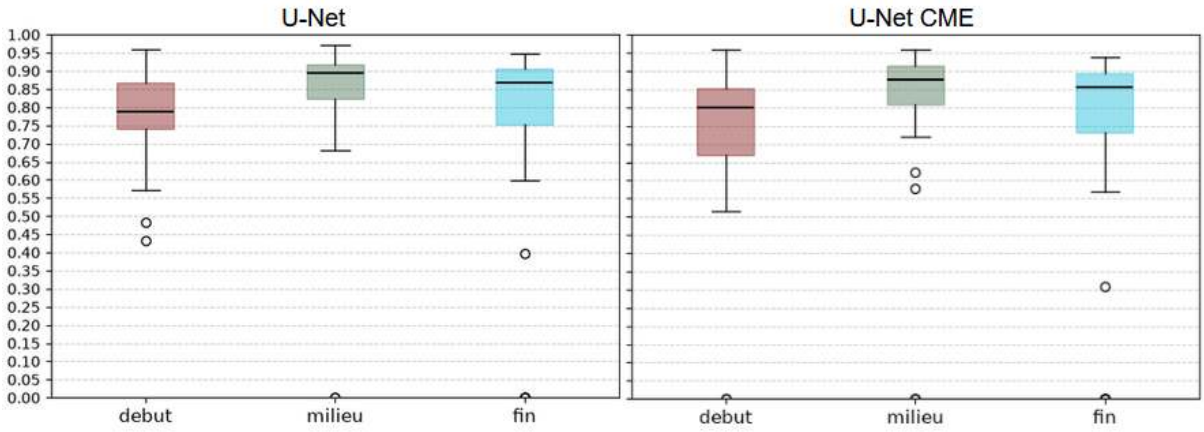


FIGURE 28 – Comparaison des scores Dice U-Net avec et sans CME pour chaque catégorie

Données : échantillon de validation.
Médiane globale : U-Net = 0.86 ; U-Net avec CME = 0.85.
Note : L'ajout du CME augmente la médiane des scores Dice pour les échographies situées au début de la gonade, mais s'accompagne également d'une plus grande variabilité. Son effet sur les valeurs extrêmes reste mitigé : certaines prédictions sont améliorées, tandis que d'autres présentent des scores Dice plus faibles.

D.3 - Fonctions de perte

La BCE (*Binary Cross Entropy*) mesure l'écart entre la probabilité prédite pour chaque pixel et son annotation de référence :

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right], \tag{3}$$

avec N le nombre de pixels considérés, y_i l'annotation de référence du pixel i et p_i la probabilité prédite par le modèle.

La Dice loss est dérivée du coefficient de Dice et mesure le défaut de recouvrement entre les masques prédits et les annotations :

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} p_i y_i + \varepsilon}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i + \varepsilon} \tag{4}$$

La MSE (*Mean Squared Error*) mesure l'écart quadratique entre les valeurs prédites et les valeurs de référence des cartes horizontales et verticales :

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}\left[(\hat{H}_i - H_i)^2 + (\hat{V}_i - V_i)^2\right], \tag{5}$$

avec H_i et V_i les valeurs de référence des cartes horizontale et verticale pour le pixel i, et $\hat{H}_i$ et $\hat{V}_i$ les valeurs prédites par le modèle.

D.4 - Hyperparamètres des modèles

TABLE 4 – Hyperparamètres des modèles

Modèle	Fct de perte	batch size	Patience scheduler	Patience arrêt
U-Net – cavité	BCE + Dice	32	3	6
ResAttentionUNet++ – cavité	BCE + Dice	16	3	6
CSA-Net – cavité	BCE + Dice	8	3	6
OpenUS – cavité	BCE	4	-	-
U-Net – œufs	BCE + Dice	32	6	15
SegHVUNet – œufs	BCE + Dice + MSE	32	6	15
YOLO26 – œufs	Composite Ultralytics	16	-	10

Pour OpenUS, l'entrainement est réalisé, comme dans l'article, sur un nombre fixe d'epochs, sans arrêt anticipé. Le checkpoint présentant le meilleur Dice sur le jeu de validation est ensuite retenu.

D.5 - Rappel et précision des modèles

TABLE 5 – Métriques de rappel et précision des modèles

Modèle	Rappel	Précision
U-Net – cavité	0.82	0.81
ResAttentionUNet++ – cavité	0.83	0.79
CSA-Net – cavité	0.79	0.72
OpenUS – cavité	0.83	0.81
U-Net – œufs	0.46	0.38
SegHVUNet – œufs	0.43	0.42
YOLO26 – œufs	0.46	0.45

Pour l'ensemble des modèles, le rappel est supérieur à la précision, ce qui signifie qu'ils ont tendance à limiter davantage les faux négatifs, au prix d'un nombre plus important de faux positifs. Cet écart est particulièrement marqué pour CSA-Net sur la cavité et U-Net sur les œufs, ce qui indique une tendance plus importante à surestimer les régions segmentées. Les autres présentent des valeurs de rappel et de précision relativement équilibrées.

E - Résultats sur l'échantillon de validation

E.1 - Dice des œufs

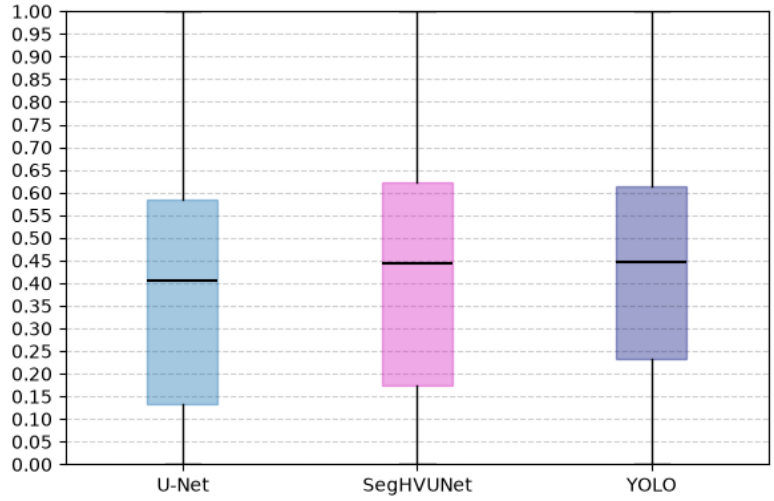


FIGURE 29 – Distribution des scores Dice des œufs

Données : échantillon de validation.
Note : Les scores Dice égaux à 1 correspondent aux échographies sans œufs pour lesquelles les modèles n'ont correctement prédit aucun œuf. Les scores Dice égaux à 0 correspondent aux échographies pour lesquelles aucun œuf en commun n'a été segmenté.

E.2 - Différence de surface des œufs

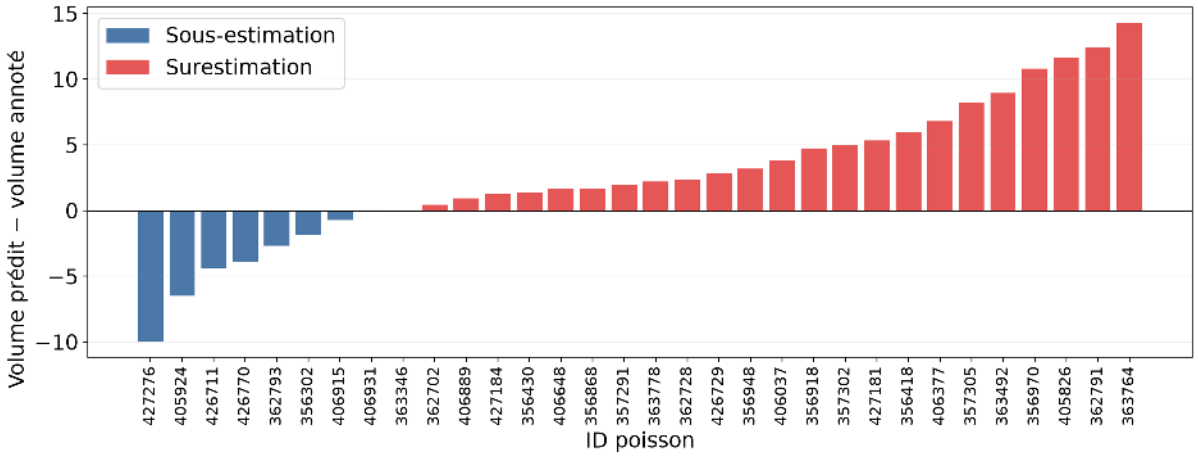


FIGURE 30 – Différence entre la surface médiane prédite et annotée des œufs avec YOLO26

Données : échantillon de validation.

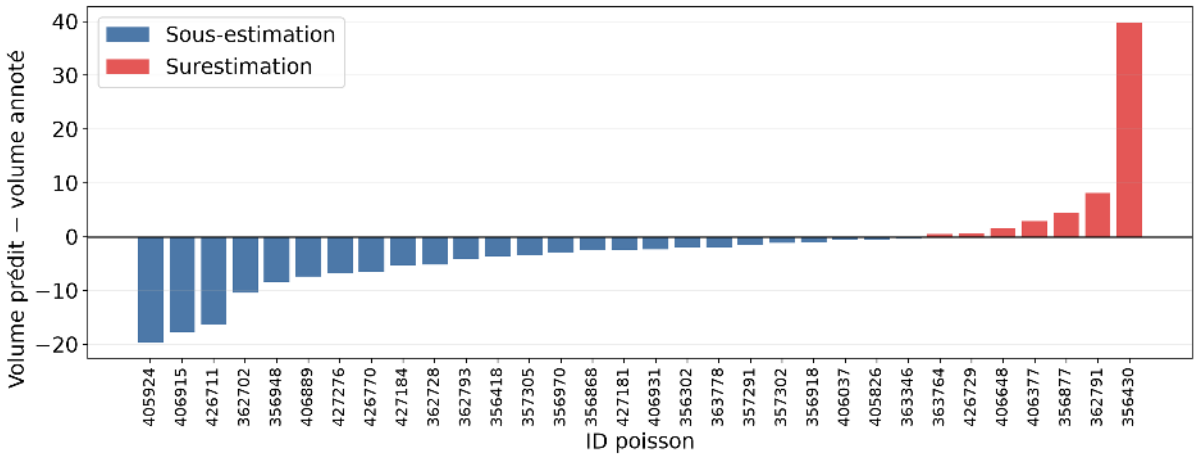


FIGURE 31 – Différence entre la surface médiane prédite et annotée des œufs avec U-Net
Données : échantillon de validation.

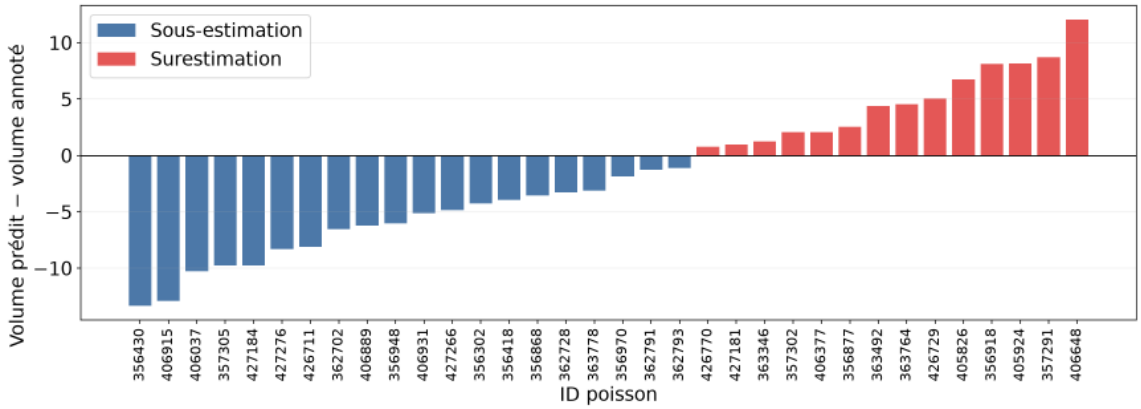


FIGURE 32 – Différence entre la surface médiane prédite et annotée des œufs avec SegHVUNet
Données : échantillon de validation.

E.3 - Différence de surface des gonades

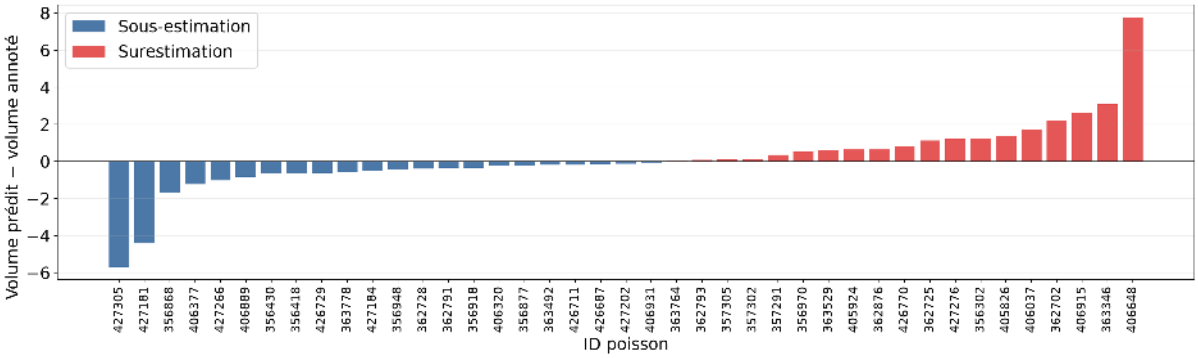


FIGURE 33 – Différence de surface entre les gonades prédites par U-Net et annotées
Données : échantillon de validation.

F - Prédictions sur de nouvelles données

F.1 - Interface

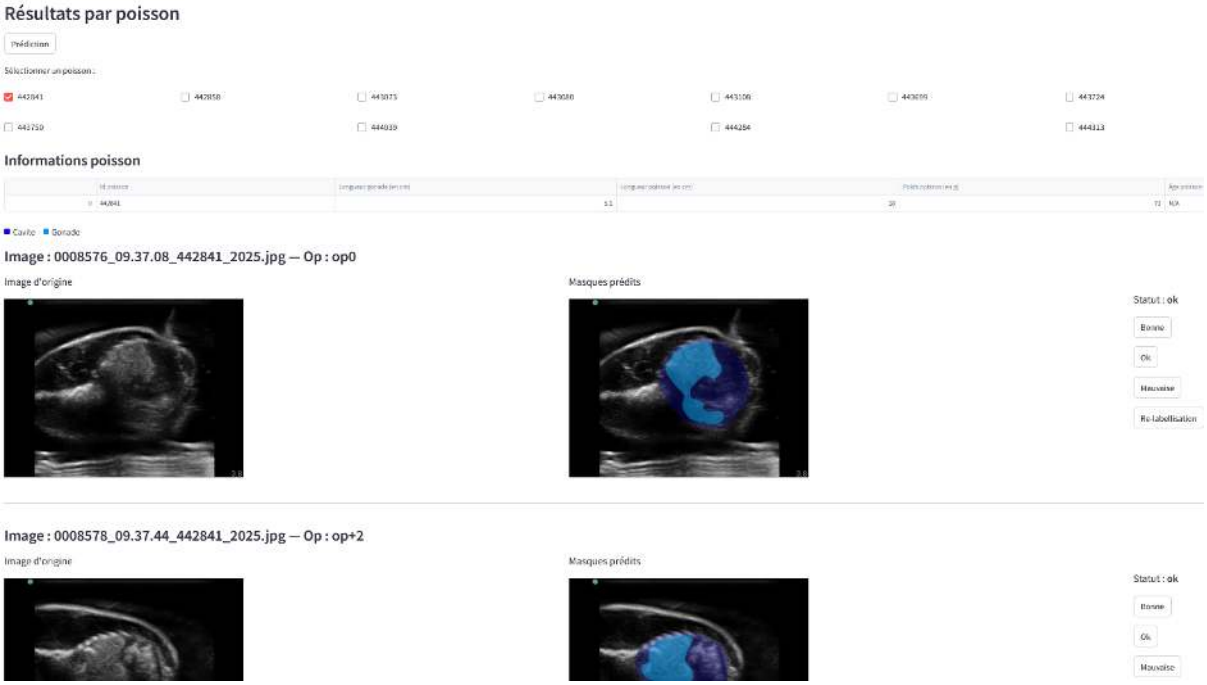


FIGURE 34 – Visualisation de l'interface sur les poissons de 2025

Note : En haut, l'utilisateur peut sélectionner un poisson parmi ceux fournis en entrée. Il obtient alors les caractéristiques du poisson ainsi que les prédictions associées, dans l'ordre, à ses différentes coupes, avec l'image d'origine correspondante. À l'aide des boutons situés à droite, il peut choisir de valider ou non la prédiction, ou de la modifier directement.

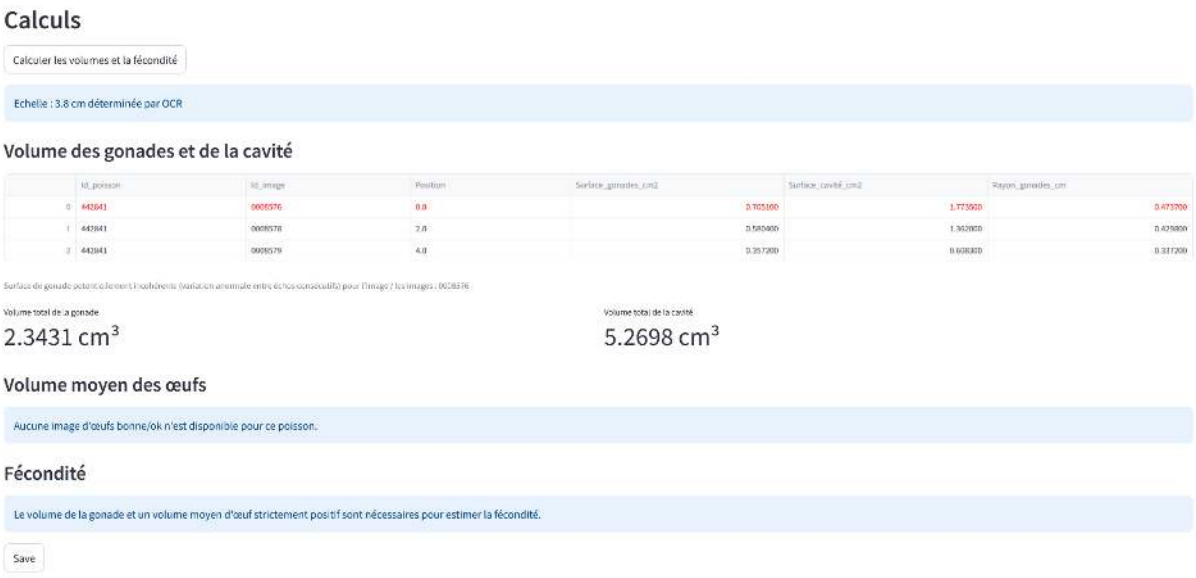


FIGURE 35 – Visualisation de l'interface sur les poissons de 2025

Note : A la fin des visualisations, l'utilisateur peut obtenir une estimation des volumes des gonades, des œufs et de la fécondité. Il peut également sauvegarder les résultats obtenus. La ligne en rouge correspond à une anomalie détectée dans l'évolution des surfaces des gonades. L'utilisateur doit ajouter les échographies longitudinales et transversales d'un même poisson pour obtenir la fécondité finale.

F.2 - Évolution des surfaces

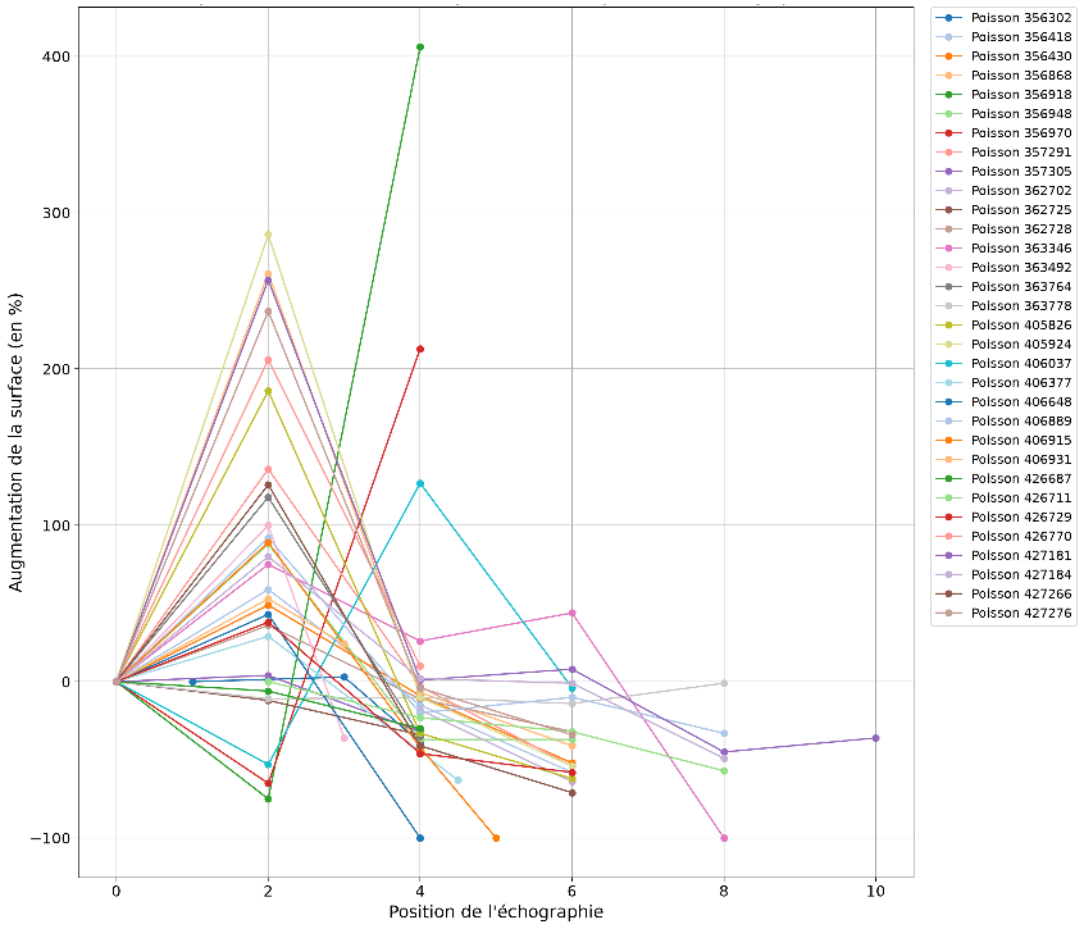


FIGURE 36 – Évolution des surfaces annotées des gonades selon la position des coupes

Note : On observe un schéma qui se répète pour la majorité des poissons, avec trois exceptions : le poisson 356918 en vert foncé, le poisson 356970 en rouge et le poisson 406037 en turquoise.

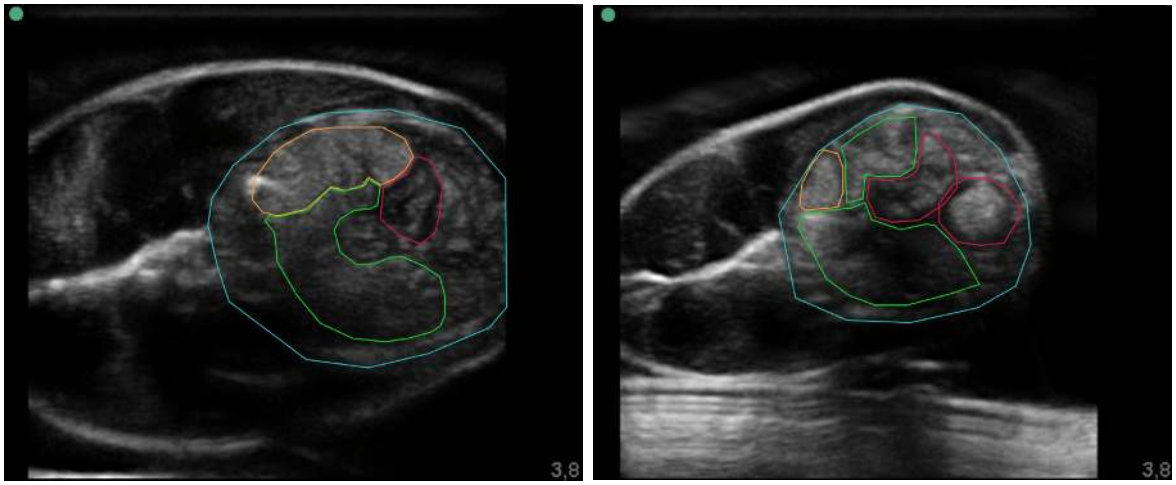


FIGURE 37 – Exemple d'échographies à la position Op+2

Note : La première échographie est celle du poisson 406037, la deuxième est celle du poisson 356918. La gonade est représentée en orange. La zone en vert correspond à la classe "Ignore gonade", utilisée lorsque l'annotateur hésitait quant à la présence d'une gonade.

G - Discussion

G.1 - Variabilité du volume des œufs

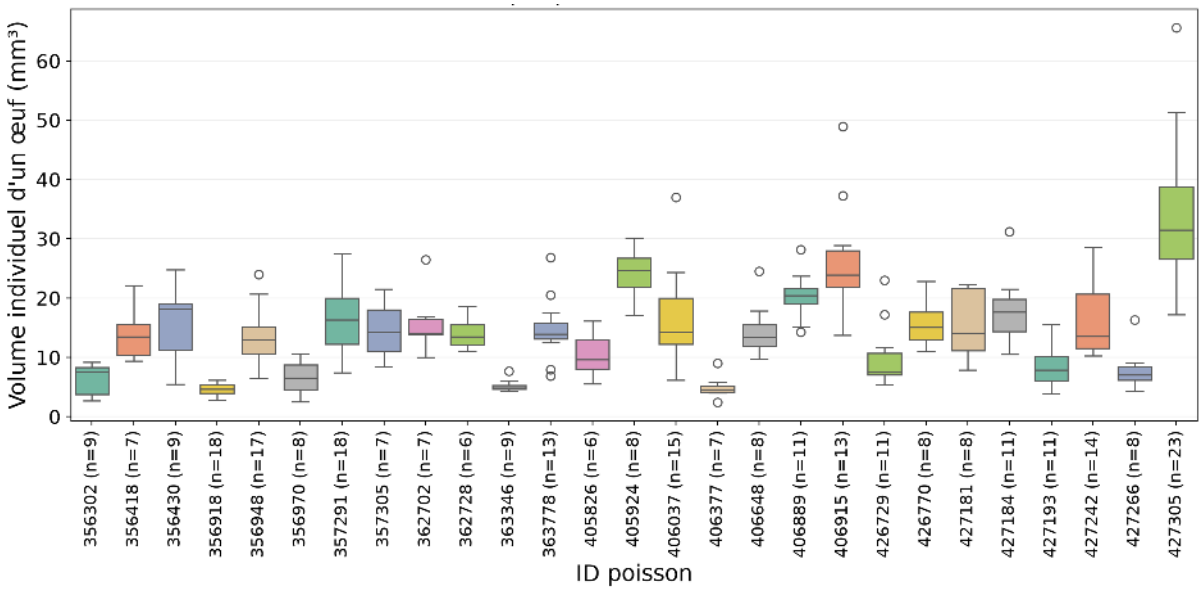


FIGURE 38 – Distribution des volumes des œufs estimés à partir des annotations

Données : poissons de l'échantillon de validation avec au moins 6 œufs annotés
Note : Pour chaque poisson, les œufs annotés sur plusieurs échographies ont été regroupés. Le nombre d'œufs annotés par poisson est indiqué. On observe une variabilité plus ou moins importante des volumes estimés entre les œufs d'un même poisson.

G.2 - Relation fécondité - longueur du poisson

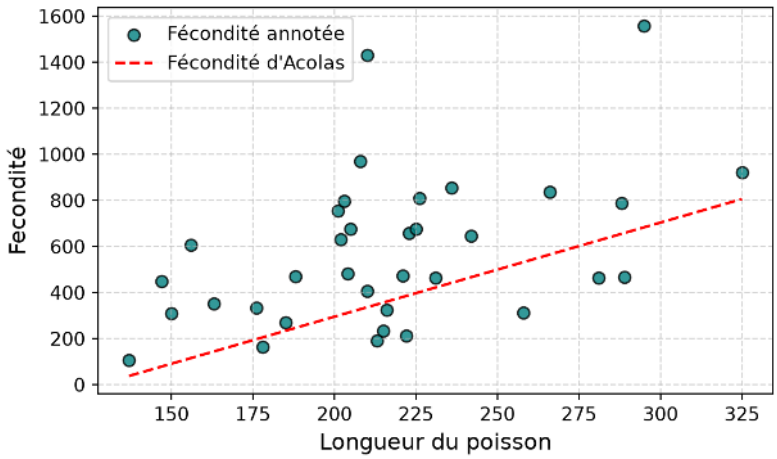


FIGURE 39 – Relation entre la fécondité estimée et la longueur du poisson mesurée

Données : fécondité calculée sur les données annotées.
Corrélation : Fécondité estimée : 0.46

Références

[1] *Roboflow*. GitHub. URL : https://github.com/roboflow.

[2] Oona RAINIO et al. « Deep Learning for Medical Ultrasound Image Segmentation : A Systematic Review of the Current Research ». In : *Journal of Imaging Informatics in Medicine* (30 mars 2026). DOI : 10.1007/s10278-026-01931-1. URL : https://doi.org/10.1007/s10278-026-01931-1.

[3] Olaf RONNEBERGER, Philipp FISCHER et Thomas BROX. *U-Net : Convolutional Networks for Biomedical Image Segmentation*. 18 mai 2015. DOI : 10.48550/arXiv.1505.04597. URL : http://arxiv.org/abs/1505.04597.

[4] Retinderdeep SINGH et al. « Advancing prenatal healthcare by explainable AI enhanced fetal ultrasound image segmentation using U-Net++ with attention mechanisms ». In : *Scientific Reports* (4 juin 2025). ISSN : 2045-2322. DOI : 10.1038/s41598-025-04631-y. URL : https://www.nature.com/articles/s41598-025-04631-y.

[5] Yasin YARI et al. « Automated Measurement of Ovary Development in Atlantic Salmon Using Deep Learning ». In : *Ultrasound in Medicine & Biology* (1er mars 2024). ISSN : 0301-5629. DOI : 10.1016/j.ultrasmedbio.2023.11.008. URL : https://www.sciencedirect.com/science/article/pii/S0301562923003691.

[6] Yasin YARI et al. « Deep Learning for Automated Egg Maturation Prediction of Atlantic Salmon Using Ultrasound Imaging ». In : *IEEE Access* (2024). ISSN : 2169-3536. DOI : 10.1109/ACCESS.2024.3409360. URL : https://ieeexplore.ieee.org/abstract/document/10547256.

[7] Alexey DOSOVITSKIY et al. *An Image is Worth 16x16 Words : Transformers for Image Recognition at Scale*. 3 juin 2021. DOI : 10.48550/arXiv.2010.11929. URL : http://arxiv.org/abs/2010.11929.

[8] Yue LI et al. *UniUltra : Interactive Parameter-Efficient SAM2 for Universal Ultrasound Segmentation*. 19 nov. 2025. DOI : 10.48550/arXiv.2511.15771. URL : http://arxiv.org/abs/2511.15771.

[9] Adrien MEYER et al. « UltraSam : A Foundation Model for Ultrasound using Large Open-Access Segmentation Datasets ». In : *International Journal of Computer Assisted Radiology and Surgery* (11 sept. 2025). ISSN : 1861-6429. DOI : 10.1007/s11548-025-03517-8. URL : http://arxiv.org/abs/2411.16222.

[10] Xiaoyu ZHENG et al. *OpenUS : A Fully Open-Source Foundation Model for Ultrasound Image Analysis via Self-Adaptive Masked Contrastive Learning*. 14 nov. 2025. DOI : 10.48550/arXiv.2511.11510. URL : http://arxiv.org/abs/2511.11510.

[11] Amarjeet KUMAR et al. *A Flexible 2.5D Medical Image Segmentation Approach with In-Slice and Cross-Slice Attention*. 30 avr. 2024. DOI : 10.48550/arXiv.2405.00130. URL : http://arxiv.org/abs/2405.00130.

[12] Juntuo WANG et al. « Conditional Metadata Embedding Data Preprocessing Method for Semantic Segmentation ». In : 2022 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC). IEEE, oct. 2022. DOI : 10.1109/CyberC55534.2022.00057. URL : https://ieeexplore.ieee.org/document/10090205/.

[13] Priyanto HIDAYATULLAH et Refdinal TUBAGUS. *YOLO26 : A Comprehensive Architecture Overview and Key Improvements*. 16 fév. 2026. DOI : 10.48550/arXiv.2602.14582. arXiv : 2602.14582[cs.CV]. URL : http://arxiv.org/abs/2602.14582.

[14] Ó THÓRISDÓTTIR et M. KIDERLEN. « Wicksell's Problem in Local Stereology ». In : *Advances in Applied Probability* 45 (déc. 2013). ISSN : 0001-8678, 1475-6064. DOI : 10.1239/aap/1386857851. URL : https://www.cambridge.org/core/journals/advances-in-applied-probability/article/wicksells-problem-in-local-stereology/1E96FD2D7FDD4F64F95A80FF7F70D186.

[15] M.-L. ACOLAS, J.-M. ROUSSEL et J.-L. BAGLINIÈRE. « Linking migratory patterns and diet to reproductive traits in female brown trout (Salmo trutta L.) by means of stable isotope analysis on ova ». In : *Ecology of Freshwater Fish* (2008). ISSN : 1600-0633. DOI : 10.1111/j.1600-0633.2007.00290.x. URL : https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0633.2007.00290.x.

[16] Qi HE et al. « Masked pretraining of U-Net for ultrasound image segmentation ». In : *Scientific Reports* (28 août 2025). ISSN : 2045-2322. DOI : 10.1038/s41598-025-11688-2. URL : https://www.nature.com/articles/s41598-025-11688-2.