Distillation de connaissances RGB-D vers RGB pour la détection de poissons dans un chalut

Stage de Master 2 / dernière année d'école d'ingénieurs

Début : février/mars 2027 (6 mois)

Lieu : Calais ou Saint-Omer.

Encadrement : Abdelbadie Belmouhcine (ULCO, LISIC) et Julien Simon (IFREMER, UMR DECOD)

Financement : gratification financée par l'EUR IFSEA, selon le barème réglementaire en vigueur

Contexte et objectifs : L'imagerie sous-marine permet d'observer les poissons dans leur environnement pour améliorer les connaissances écologiques des zones côtières [1]. Utilisée directement dans les engins de pêche, elle permet également de recueillir des informations utiles à l'étude des ressources halieutiques. De nouveaux cas d'usage sont également apparus pour améliorer la sélectivité (capacité à garder uniquement les espèces ciblées par les pêcheurs) des engins de pêche [2]. La détection et la reconnaissance automatiques des espèces restent toutefois difficiles en raison des conditions de visibilité, des mouvements des poissons et des variations de leur distance à la caméra. Cette distance influence la taille apparente des poissons sur les images et la visibilité des caractéristiques permettant de les identifier. Sur les poissons proches de la caméra, certains détails peuvent faciliter la distinction entre espèces, comme le maquereau, le chinchard ou la sardine. Pour les poissons plus éloignés de la caméra, ces indices deviennent moins discernables et les confusions de classe peuvent augmenter. Un réseau utilisant uniquement l'image couleur (RGB) doit apprendre à gérer ces variations sans disposer explicitement de la distance ni de la taille physique des individus. Les cartes de profondeur obtenues par une caméra stéréo permettent d'apporter des informations géométriques complémentaires. Elles peuvent notamment servir à estimer la distance à la caméra et, sous réserve d'une calibration et d'une précision suffisantes, la taille physique des poissons. L'hypothèse du stage est que ces informations peuvent aider un réseau à adapter l'exploitation des caractéristiques visuelles aux conditions d'observation.

Le stage aura deux objectifs : développer un modèle enseignant exploitant conjointement l'image et la profondeur, puis transférer les connaissances acquises à un modèle étudiant utilisant uniquement une image RGB 2D. Ce transfert sera réalisé par distillation [1], afin d'évaluer quelle part du bénéfice de l'apprentissage avec profondeur peut être conservée sans entrée stéréo ni calcul de profondeur à l'inférence.

[1] https://kosmos.fish/
[2] https://gameoftrawls.IFREMER.fr/

Travail proposé : Le travail s'appuiera sur les jeux de données stéréo annotés disponibles à IFREMER, acquis avec une caméra ZED Mini [3]. De nouvelles acquisitions avec une ZED X Mini pourront compléter ces données. Les cartes de profondeur seront générées à partir des images stéréo avec les outils disponibles. La calibration et les espèces ciblées seront précisées avec IFREMER. Le périmètre sera limité à quelques catégories pertinentes et à une architecture existante de détection. Une évaluation complémentaire sur un benchmark public RGB-D, tel que SUN RGB-D [2], pourra être menée.

Le stage sera organisé en trois étapes :

1. État de l'art, préparation des données et modèle de référence
    - Étude des méthodes de reconnaissance sous-marine, d'apprentissage associant image et profondeur et de distillation de connaissances.
    - Vérification de la calibration, de la qualité des cartes de profondeur et des annotations disponibles.
    - Séparation des séquences d'acquisition entre apprentissage, validation et test afin de limiter les biais liés à la proximité entre images.
    - Évaluation des performances d'un modèle de référence utilisant uniquement l'image RGB 2D.
2. Apprentissage d'un enseignant utilisant l'image et la profondeur
    - Développement d'un modèle exploitant conjointement les caractéristiques visuelles et les informations de profondeur.
    - Étude d'un mécanisme permettant de conditionner leur pondération par la distance à la caméra et, lorsque la fiabilité des mesures le permet, par la taille physique estimée des poissons. Cette taille sera calculée à partir de la géométrie stéréo et de la localisation des individus.
3. Distillation vers un modèle sur image RGB 2D
    - Entraînement d'un modèle étudiant recevant uniquement l'image 2D, supervisé à la fois par les annotations et par les prédictions de l'enseignant. Le transfert portera d'abord sur les sorties du modèle. Une distillation des caractéristiques intermédiaires sera étudiée selon l'avancement. La comparaison d'un même étudiant entraîné avec et sans distillation permettra d'isoler le bénéfice du transfert.

L'évaluation portera sur le mAP global et par espèce. Les résultats seront également analysés selon la distance à la caméra et la taille apparente des poissons. Des études d'ablation distingueront l'apport de la profondeur, des indicateurs de distance et de taille, puis de la distillation. L'étudiant sera comparé à la référence RGB sans distillation sur des séquences de test ZED Mini et, selon les données disponibles, sur un autre capteur (GoPro ou I2S), afin d'évaluer le gain et sa robustesse au changement de caméra. Le temps d'inférence, le nombre de paramètres et le nombre d'opérations par image, exprimé en GFLOPs, seront également mesurés pour le modèle étudiant.

[3] https://www.stereolabs.com/en-fr/developers

Profil recherché : Étudiant ou étudiante en Master 2 ou dernière année d'école d'ingénieurs, en informatique, intelligence artificielle, traitement d'images ou domaine proche.

Compétences attendues :

- Bonnes connaissances en apprentissage automatique et apprentissage profond.
- Maîtrise de Python et pratique d'un framework tel que PyTorch.
- Intérêt pour la vision par ordinateur et l'expérimentation scientifique.
- Rigueur, autonomie et capacité à analyser et présenter des résultats.
- Capacité à lire des publications scientifiques en anglais.
- Des connaissances en détection d'objets, vision stéréo ou distillation seraient appréciées, mais ne sont pas requises.

Environnement scientifique : Le ou la stagiaire sera encadré(e) par Abdelbadie Belmouhcine, maître de conférences au Laboratoire d'Informatique, Signal et Image de la Côte d'Opale (LISIC) de l'Université du Littoral Côte d'Opale (ULCO), et Julien Simon, IFREMER, UMR DECOD. Le stage bénéficie d'un financement de l'EUR IFSEA. L'accueil se fera au LISIC, sur le site de Calais ou de Saint-Omer.

Candidature : Envoyer un CV, les relevés de notes récents et quelques lignes présentant votre intérêt pour le sujet à : Abdelbadie Belmouhcine - `abdelbadie.belmouhcine@univ-littoral.fr`

Références

[1] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[2] Shuran Song, Samuel P Lichtenberg, and Jianxiong Xiao. SUN RGB-D: A RGB-D scene understanding benchmark suite. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 567–576, 2015.