Proposition de thèse CIFRE

Apprentissage frugal et compressé pour l'interprétation des scènes sous-marines

ISEN Yncréa Ouest/Thales - Site de Brest

Encadrement

Directeur de thèse : Maher JRIDI, Enseignant chercheur HDR, Équipe Vision-AD - L@bISEN
Encadrants :

- Thibault NAPOLÉON, Enseignant chercheur, Équipe Vision-AD - L@bISEN
- Ayoub KARINE, Enseignant chercheur, Équipe Vision-AD - L@bISEN
- Franck FLORIN, Thales

Sujet

L'océan couvre la plus grande partie de la surface de la Terre (environ 71%) avec une superficie de 361 millions de kilomètres carrés. De ce fait, cette richesse naturelle est étudiée avec la plus grande attention pour répondre à des enjeux écologiques et économiques cruciaux. Dans ce sens, les recherches scientifiques peuvent se scinder en deux grandes familles à savoir l'étude de la surface des océans et l'étude du milieu sous-marin. C'est dans le contexte général de l'interprétation du milieu sous-marin que s'articule le présent sujet de thèse. L'interprétation directe de ce milieu par un être humain reste une tâche risquée, coûteuse et difficile compte tenu du temps et des coûts liés aux systèmes d'acquisition et aux missions sous-marines. En conséquence, il est nécessaire de développer de nouvelles méthodes visant à définir des outils technologiques de prise de décision pour l'exploration automatique du milieu sous-marin.

Les données concernant le milieu sous-marin peuvent être acquises via des systèmes d'observation à l'instar des drones sous-marins autonomes (AUV, Autonomous Underwater Vehicle). Ces systèmes sont dotés de capteurs de vision leur permettant d'acquérir des vidéos sous-marines optiques. L'analyse automatique de ces flux vidéo permet d'interpréter avec plus de finesse l'espace sous-marin pour des missions de cartographie, d'études des habitats, de suivis de structures sous-marines ou encore pour la recherche d'objets immergés. Dans la présente thèse, nous nous intéressons à la segmentation sémantique qui vise à affecter une classe à chaque pixel dans les vidéos. Ainsi, la carte de segmentation générée représente une classification fine des différentes zones (substrats et objets) de la scène observée. En vision sous-marine, cette tâche est effectuée en s'inspirant des méthodes d'intelligence artificielle ayant montré leurs suprématies dans l'interprétation du milieu aérien à l'instar des réseaux de neurones profonds [1, 2]. Étant donné qu'une vidéo est une succession d'images (frames), la solution la plus directe pour sa segmentation sémantique est d'appliquer un modèle de prédiction à chacune de ses images en utilisant un CNN (Convolutional Neural Network) par exemple [3]. Cependant, cette solution naïve ne fait pas preuve de bonne performance pour la segmentation. Ceci est dû principalement à la non-considération de la relation temporelle entre les images de la vidéo. Contrairement aux images statiques, l'information temporelle est d'une grande importance dans le traitement des vidéos. Elle permet de modéliser la progression de la scène observée dans le temps. Par conséquent, les travaux récents traitant cette problématique essaient de prédire les classes associées aux pixels

d'une image à l'instant "t" en utilisant les classes affectées aux pixels des images précédentes ("t-1", "t-2", etc.). Pour ce faire, plusieurs travaux, après la segmentation sémantique de chaque image, ajoutent un module d'agrégation (flux optique, tracking, etc.) suivi d'un réseau de neurones séquentiel (RNN, LSTM, Transformer, etc.). D'autres familles de méthodes n'utilisent que quelques images de la vidéo (appelés keyframes) et propagent les cartes caractéristiques vers les autres images à travers le flux optique. Néanmoins, l'adaptation de ces méthodes en vision sous-marine a montré des limites en termes de robustesse, mais aussi en termes de temps de calcul. Ces deux limites sont respectivement liées, principalement, à deux facteurs : (1) la non-disponibilité d'un grand jeu de données étiquetées de vidéos sous-marines, ceci à cause des coûts élevés des missions d'acquisition (systèmes coûteux et annotation manuelle chronophage des vidéos pour les experts). (2) le nombre important des opérations et des paramètres utilisés dans les approches neuronales pour la segmentation des vidéos sous-marines. Pour combler ces deux limites, le présent sujet de thèse vise à proposer des architectures neuronales compressées capables d'apprendre à partir d'un très faible volume de vidéos et d'offrir des performances de segmentation élevées.

Concernant la première limite, la piste de l'apprentissage frugal (Few-Shot Learning) [4, 5] sera étudiée. Il s'agit d'une approche apte à apprendre à partir d'un nombre limité de données d'apprentissage étiquetées. Cela est atteint en se basant sur l'accumulation de différentes connaissances préalables extraites à partir d'autres bases de données (appelée donnée de base) plus grandes. Cette étape fait référence à l'apprentissage de représentation. Finalement, le nombre réduit des données d'apprentissage (appelées support) est utilisé pour construire la fonction de décision.

Quant à la deuxième limite, la distillation de connaissances (Knowledge Distillation) [6, 7] pourrait être une approche prometteuse. Le principe de cette approche consiste à transmettre les connaissances d'un grand réseau de neurones qui donne de bonnes performances sur une tâche spécifique, dit enseignant, vers un autre réseau réduit, dit étudiant. L'objectif est que le réseau étudiant imite l'apprentissage du réseau enseignant. Autrement dit, l'apprentissage du réseau étudiant est supervisé par le réseau enseignant. Ainsi, il est possible d'aboutir à un réseau de neurones performant en segmentation sémantique avec une taille réduite, facilement embarquable dans les drones autonomes sous-marins grâce à sa rapidité d'inférence et sa consommation énergétique réduite. La méthode mise en œuvre pourra ainsi être utilisée pour segmenter, en temps réel et précisément, les fonds marins en fonction des substrats présents. Enfin, des données dynamiques 3D, de type nuage de points, pourront être utilisées afin de renforcer la segmentation sémantique et produire des relevés encore plus précis pour créer une cartographie des fonds marins.

Compétences attendues

Dans l'idéal, le candidat doit avoir :

- suivi un cursus de Master ou d'Ingénieur dans un des domaines suivants : intelligence artificielle, vision par ordinateur, science des données, mathématiques appliquées ;
- de solides compétences en algorithmique et en programmation : Python, PyTorch, Tensoflow, Keras... ;
- des connaissances en apprentissage profond appliqué à la vision par ordinateur.

Modalités de la thèse

- **Type de contrat :** thèse CIFRE sous contrat avec Thales
- **Laboratoire d'accueil :** équipe Vision-AD du L@bISEN
- **Lieu de la thèse :** Brest
- **Nationalité :** Le candidat doit obligatoirement avoir la nationalité française

Candidature

Le candidat doit envoyer un email, dont le sujet sera `[TheseVisionSousMarine]`, aux personnes suivantes :

- maher.jridi@isen-ouest.yncrea.fr
- thibault.napoleon@isen-ouest.yncrea.fr
- ayoub.karine@isen-ouest.yncrea.fr

Cet email devra contenir dans l'idéal :

- un CV détaillé ;
- une lettre de motivation ;
- des lettres de recommandation (professeurs, encadrant de stage...) ;
- les relevées de notes des deux dernières années (S7, S8 et S9).

Les candidatures seront étudiées au fil de l'eau.

Références

[1] Islam, M. J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., & Sattar, J. "Semantic segmentation of underwater imagery : Dataset and benchmark", IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020.

[2] Chicchon, M., & Bedon, H. "Semantic Segmentation of Underwater Environments Using DeepLabv3+ and Transfer Learning", Smart Trends in Computing and Communications (pp. 301-309). Springer, Singapore, 2022.

[3] T. Zhou, F. Porikli, D. J. Crandall, L. V. Gool and W. Wang, "A Survey on Deep Learning Technique for Video Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.

[4] Snell, J., Swersky, K., & Zemel, R. "Prototypical networks for few-shot learning", Advances in neural information processing systems, 30, 2017.

[5] Wang, K., Liew, J. H., Zou, Y., Zhou, D., & Feng, J. "Panet : Few-shot image semantic segmentation with prototype alignment", Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9197-9206), 2019.

[6] Liu, Y., Shu, C., Wang, J., & Shen, C. "Structured knowledge distillation for dense prediction", IEEE transactions on pattern analysis and machine intelligence, 2020.

[7] Karine, A., Napoléon, T., Jridi, M., "Semantic Images Segmentation for autonomous driving using Self-Attention Knowledge Distillation", 16th IEEE International Conference on Signal Image Technology & Internet Based Systems, Oct 2022, Dijon, France, 2022.