

STAGE M2

Extraction et reconnaissance automatique sur des photos d'invertébrés marins dans un contexte de cartographie d'écosystèmes.

Contexte Le Muséum national d'Histoire naturelle assure le suivi scientifique des pêcheries australes françaises, notamment en tant qu'expert institutionnel sur la biodiversité auprès de l'Etat. Cette mission comprend la fourniture d'avis scientifiques sur l'état des peuplements d'invertébrés marins impactés par la pêche française dans des territoires et eaux internationales de l'océan Austral. La construction de modèles spatiaux permet de mettre en évidence les patrons de distribution des différentes espèces et de caractériser leur niche écologique [1]. L'étude des assemblages d'espèces permet de cartographier les écosystèmes marins benthiques potentiellement impactés par la pêche [2].

Données Ces travaux de modélisation reposent sur l'utilisation de données produites qui incluent notamment l'échantillonnage photographique de tous les invertébrés marins (p.ex. coraux, éponges, anémones de mer, étoiles de mer, oursins) collectés par les engins de pêche (palangre, chalut). A partir de l'identification des espèces visibles sur ces photos, du dénombrement des organismes et de leur mesure, il est possible d'établir des distributions d'abondances qui alimentent les modèles permettant la cartographie des écosystèmes marins vulnérables. Le procédé actuel consiste à extraire manuellement par vignettage des images. Ces vignettes sont ensuite identifiées par des naturalistes, au niveau de l'espèce ou au niveau d'un groupe d'espèces. Une base de données de près de 90 000 photos d'invertébrés marins de l'océan Austral, principalement de Kerguelen, identifiées et classées est actuellement disponible.

Objectif L'objectif du stage est de développer un outil d'analyse automatique des photographies collectées qui permette de localiser et d'identifier les différentes espèces. Dans un contexte où les algorithmes de détection d'objets ont récemment permis de notables améliorations des performances ([3]), on pourra procéder par transfert d'apprentissage ([4,5]), de manière à identifier la meilleure manière d'adapter des algorithmes dont l'apprentissage a été effectué sur des bases de photographies génériques au contexte particulier des données considérées (conditions d'acquisition peu contrôlées, présence systématique d'occultations, similarité d'aspect entre espèces différentes). Par ailleurs, le volume conséquent de données disponibles devrait permettre de comparer les résultats du transfert d'apprentissage à une approche reposant sur un apprentissage « end-to-end ». Dans un deuxième temps, le stage peut évoluer vers le développement d'outils d'analyse des données longitudinales de distribution des espèces.

Encadrement

Alexis Martin, Muséum National d'Histoire Naturelle, alexis.martin@mnhn.fr

Yann Gousseau, Télécom ParisTech, yann.gousseau@telecom-paristech.fr

Bibliographie

[1] Guillaumot, C., Martin, A., Eléaume, M., Saucède, T..Methods for improving species distribution models in data-poor areas: example of sub-Antarctic benthic species on the Kerguelen Plateau. *Marine Ecology Progress Series*, (594), 149–164, 2018

[2] Martin, A., Trouslard, E., Hautecoeur, M., Blettery, J., Moreau, C., Améziane, N., Saucède, T., Duhamel, G., Eléaume, M. *CCAMLR's Vulnerable Marine Ecosystems*

bioindicator taxa : a relevant tool for benthic ecoregionalisation, CCAMLR Working Group on Ecosystem Management and Monitoring, 18/19, 17pp, 2018

[3] Ren, S., He, K., Girshick, R., & Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6), 1137-1149, 2017

[4] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. *ICML 2014*, pp. 647-655

[5] Huh, M., Agrawal, P., & Efros, A. A. What makes ImageNet good for transfer learning?. *arXiv preprint*, 2016

Fig.1: description of the pictures treatment pipeline: count and identification of the organisms (A), extraction of the cropped pictures of each organism (B), measurement (C), storage into a database (D) ready to be extracted for analysis and statistics

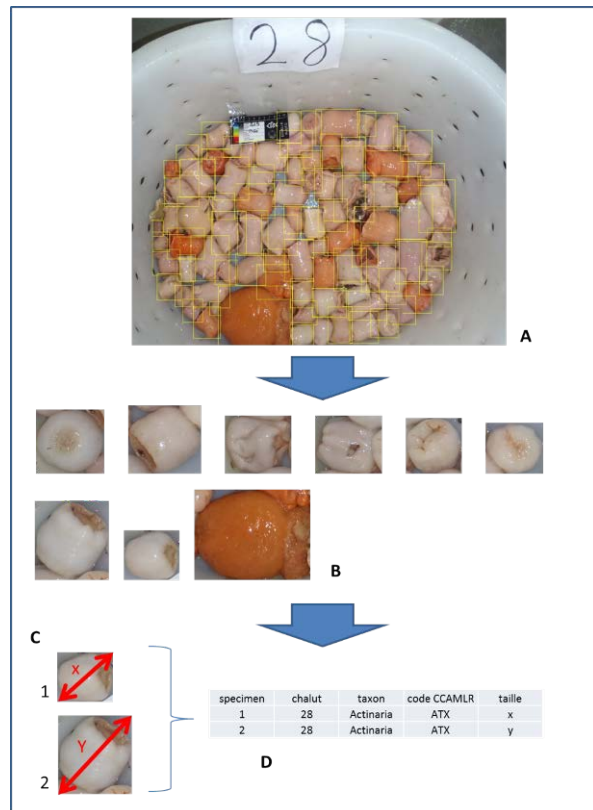


Fig.2: a typical photograph, from which organisms need to be automatically localized and classified

